

# LOS ALAMOS NATIONAL LABORATORY

## **Detection and Location of Radioactive Sources using a Suite of Slab Detectors**

James W. Howse  
Computational Science Methods Group, X-CM  
Mail Stop F645

LANL Technical Report: LA-UR-99-908

Report Date: February 24, 1999



## Abstract

This report describes several experiments used to characterize and test a suite of four radiation sensors. The sensors are scintillation counters composed of plastic connected to an amplifier. The purpose of these tests is to assess the feasibility of using these sensors to detect and track radioactive sources in a large room. The tests are also used to compare a number of different detection and tracking algorithms.

We describe real-time algorithms for both detecting the presence and tracking the position of radioactive sources in a facility in the presence of measurement noise. We formulate the detection problem as a nonparametric hypothesis testing problem. This problem is solved by comparing a statistic computed over some window(s) of the data to a threshold value. If this threshold is exceeded then we decide that a source is present. We formulate the tracking problem as a state estimation problem and solve it recursively using a constrained nonlinear optimization method. The optimization simultaneously minimizes the change in source position and disagreement between measurements and a sensor model. The sensor model is a fairly complex function relating position to detected count rate.

The overall purpose of this work is to enhance both security and safety by automating part of the assessment process, allowing remote assessment, and introducing new sensor modalities into the assessment process. We present detection and tracking results based on experiments done with one source in a single room. Our results indicate that a source can be detected and tracked quite well with these algorithms in spite of fairly poor signal to noise ratios, and rather high measurement noise levels. In short, we demonstrate the capability to detect and track a single source in real-time with high accuracy in spite of a complex mapping from source position to detected count rate, an unknown background signal, and high measurement noise.

## Key Words

Radioactive Source Tracking and Detection, Nonparametric Hypothesis Testing, Nonlinear Least Squares Estimation

## Acknowledgments

We would like to thank Larry Ticknor, Ken Muske, Kevin Buescher, and Don Hush for valuable insights and enlightening discussions in the course of this work.

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.



# 1 Introduction

In this report we discuss various aspects of the problem associated with monitoring radioactive sources in a particular facility. The facility in question conducts various experiments using radioactive materials. Knowledge of the source locations has implications for both security and safety. Currently all security and safety assessments are made by people who are physically present during experiments. The overall purpose of the project is to provide tools to facilitate and enhance this assessment process. Our algorithms are one part of a suite of tools which will allow some of these assessments to be made remotely. The tools will also provide sensor modalities that are not available to people using only their own senses. These tools will also be used to partially automate the assessment process. These three factors will enhance both security and safety by reducing personnel risk through remote assessment, providing new methods and sensor modalities for risk assessment, and providing a source of independent verification for the current assessment process.

This report consists of four major sections. The first section describes an experiment that was performed in order to characterize the radiation sensors. The second section consists of conclusions about the sensor characteristics based on an analysis of this data. The third section describes several algorithms for detecting the presence of radioactive sources using the given sensors. The fourth section discusses several procedures for tracking moving radioactive sources with these sensors.

The data for this problem is four time series consisting of the count rate at one second intervals from four gamma ray detectors which are located at four different positions within the room. The count rate at a particular sensor is the total number of gamma-energy photons received by the sensor during a one second time interval. The specific sensors that we use consist of a photoluminescent slab of plastic attached to a photomultiplier tube. The predominant source of measurement noise comes from the stochastic nature of gamma emissions from the source itself. Ideally, the emission of gamma photons from a radioactive source over time looks like a series of samples drawn from a Poisson distribution. Since the variance of a Poisson distribution is equal to its mean, the deviation of the detector readings is proportional to the square root of the mean count rate.

For source detection, we describe a number of nonparametric detection strategies for determining the presence of radioactive sources in a facility. We detect a source by comparing a statistic computed over some window(s) of the data to a threshold value. If the statistic value exceeds the threshold value, then we decide that a source is present. All of these detection methods are forms of nonparametric hypothesis testing. More specifically they fall into two broad categories. The first category computes a statistic over some window of the time series from the sensors, and reports that a source is present when the value of this statistic exceeds some threshold. The second category splits this window in half, computes a statistic in each window half and compares the similarity of these two statistics. If the two statistics are sufficiently different, then the method reports that a source is present. Any detection method must be able to measure a change in a noisy time series signal against an unknown background signal. The signal noise results from the signal being a Poisson process, and the background uncertainty arises from being unable to measure the background level directly when a radioactive source is present. The best source detection algorithm that we investigated from the first category of methods used the variance to mean ratio as the statistic. The best strategy that we tested from the second category used the Kolmogorov-Smirnov statistic to compare the two half windows.

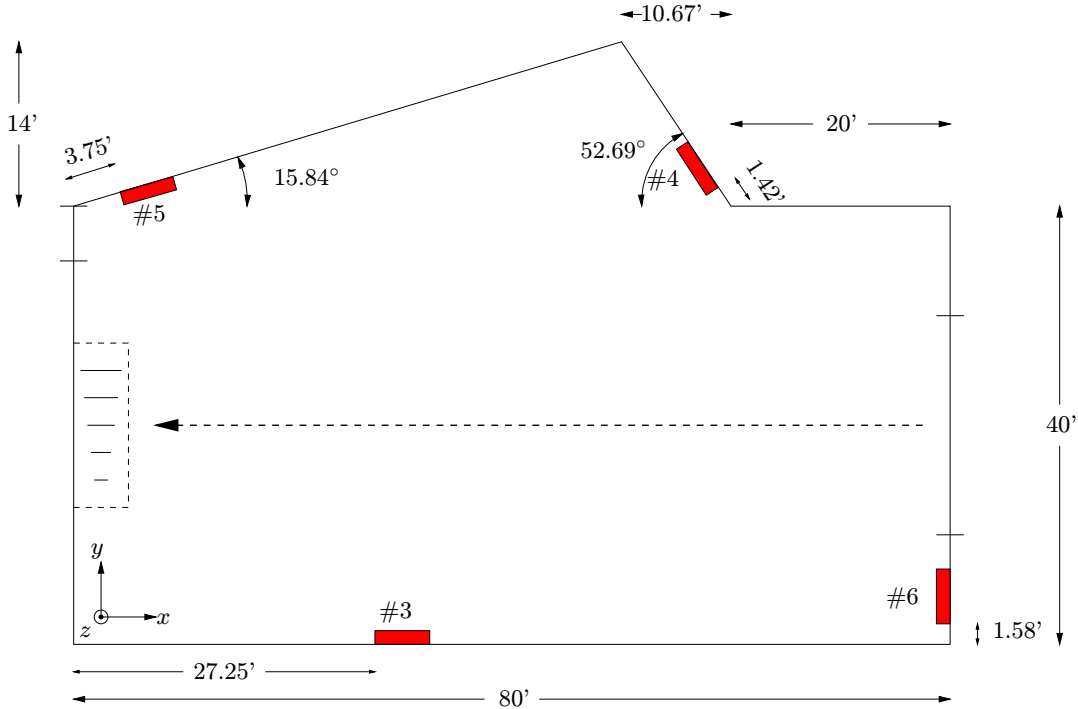
For source tracking, we describe both recursive and moving horizon algorithms for recursively estimating the real-time positions of radioactive sources in a facility. In this work, we will explicitly consider the problem of tracking a single source within one room. We estimate the position by assuming a known initial source position and then estimating the change in its position from this initial state. Our data consists of the measured time series from the four sensors described previously. Note that these sensors *do not* form an image of the room in the way that a camera would. This means that many of the techniques for locating moving objects in images can not be applied. Our approximations of the detected count rate are based on a nonlinear model of each sensor which relates the source lo-

cation to the measured count rate at that sensor. The best source tracking algorithm that we tried was a recursive nonlinear least squares estimator which simultaneously minimizes the expected value of the difference between the detector measurements and the model predicted count rates, and also the expected value of the change in the source location, in the presence of noise. Because the measurements come from a Poisson process, the uncertainty in the source location based on the detector readings is roughly proportional to the square root of the mean count rate.

An analogy may clarify the difficulties associated with the tracking problem. Imagine a building containing only one room and having a flat roof with four skylights cut into the ceiling, each near one corner of the building. You are standing on the roof and someone is walking around in the room holding a candle. You must determine the position of the candle in the room by observing the relative brightness of the light coming through the four skylights. Keep in mind that the flickering of the candle leads to variations in its brightness that are proportional to the square root of the brightness itself. This analogy makes it clear that many image processing techniques for tracking moving objects would be ineffective on this problem.

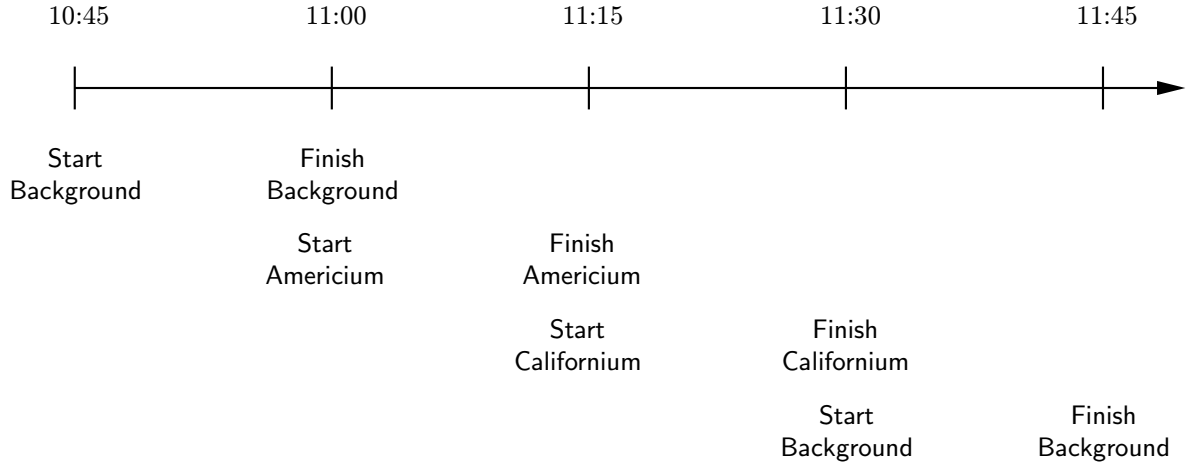
## 2 Description of Experiment

The experiment took place in the high-bay of TA-18. The layout of this room is illustrated in Figure 1. The four radiation sensors are denoted by the gray rectangles lying against the walls. They are labeled



**Figure 1:** A diagram of the layout of the facility. The thick dotted line running through the middle of the room indicates the path that the source was walked along.

by the number of the communications port used to report their data to the computer system. All distances are measured in feet and all angles in degrees. This experiment followed the time line shown in Figure 2. The Start Background – Finish Background phase consisted of taking readings of the radiation level in the high-bay with no other radioactive sources present. During the Start Americium – Finish



**Figure 2:** The time line of the characterization experiment.

Americium phase an *Americium Fluorine* source whose tested strength was  $\sim 10^5$  was placed in the room at 16 different locations for approximately 45–60 seconds at each location. During the Start Californium – Finish Californium phase a *Californium* source whose tested strength was  $\sim 10^6$  was placed in the room at these same positions for the same time duration at each of the positions. Conceptually the source testing phases (i.e. Start Americium – Finish Americium and Start Californium – Finish Californium) each consist of three steps.

**Step 1:**

Hold the source at 15 locations clustered around Sensor #3 for 45–60 seconds apiece. The position of each test point was measured with respect to the center of this sensor. The purpose of this step was to take data which will allow the sensitivity with respect to distance and angle to be characterized.

Note that while testing the *Americium* source, the steel cover of the sensor was raised and held up for 45–60 seconds while the source was at positions 1 and 7 in Figure 3. This was done immediately following the regular measurements at these locations.

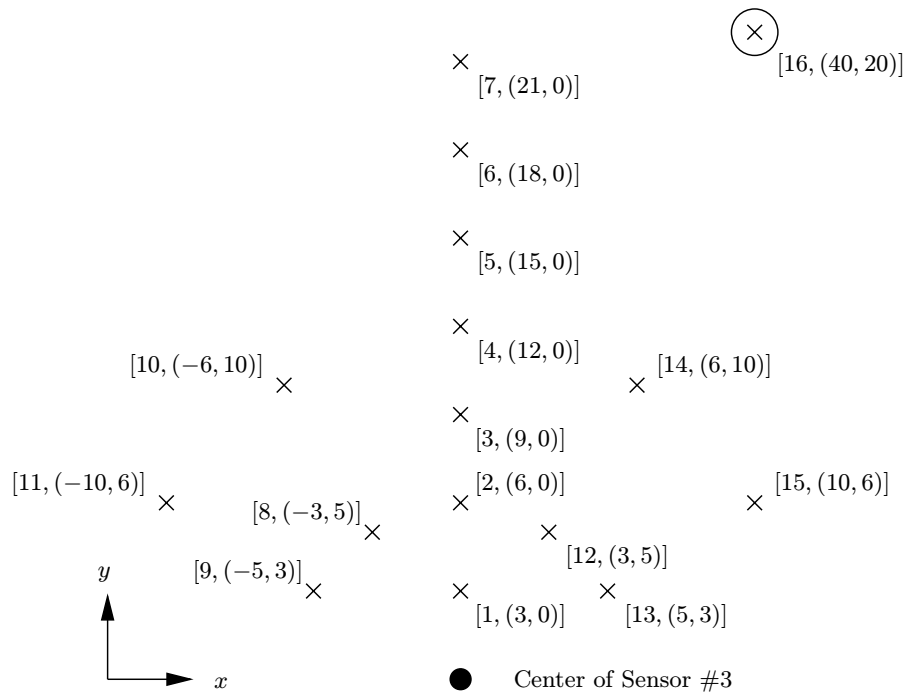
**Step 2:**

Hold the source at the center of the room for approximately 60 seconds. This step provides data to assess whether the reading from each sensor is higher than the background reading when the source is at its maximum distance from all sensors.

**Step 3:**

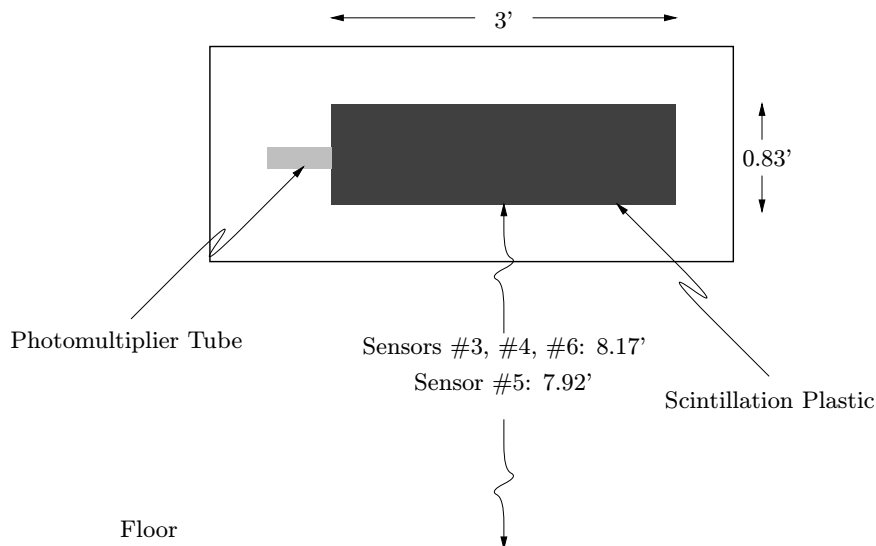
Walk the source down the middle of the room in the long direction. The data from this will be used to test the algorithms which try to identify the source location from sensor data.

The room locations at which measurements were made are shown in Figure 3. Each  $\times$  in this figure indicates the location of a sensor measurement. The numbers beside each  $\times$  are [ *measurement number*, ( *x location*, *y location* ) ]. The *measurement number* denotes the order in which measurements were taken for each source, in other words 1 is the first measurement, 10 the tenth measurement, etc. The pair ( *x location*, *y location* ) specify the distance in feet along the *x* and *y* axis respectively from the center of the sensor at position  $\bullet$  to the measurement location at  $\times$ . So the pair  $(-5, 3)$  indicates that the measurement was taken at a point 5 feet to the left of the sensor (from the sensor's perspective) and 3 feet back from it. Sensor #3 is located in the middle of the wall lying to the right as the high-bay is entered. The circled  $\times$  is located approximately at the center of the room and the ( *x location*, *y location* ) pair for this point gives the distance from the walls in the *x* and *y* directions (i.e. the lower left hand corner of Figure 1). Following the readings at these 16 locations, the source was walked down



**Figure 3:** The measurement locations for the characterization experiment.

the middle of the room as shown by the heavy dotted arrow in Figure 1. Information about the sensor size and vertical position is shown in Figure 4. Note that the sensor height is measured from the floor to



**Figure 4:** The size and vertical position of the radiation sensors.

the bottom of the plastic material inside the steel housing, *not* to the bottom of the steel housing. Also note that Sensor #5 is slightly lower than the other three sensors. The *efficiency* of a detector is defined as the ratio of the number of pulses recorded by the detector to the number of gamma photons incident on the detector. For these detectors, the efficiency is rated at  $\sim 10\%$  when tested with a Cesium-137

source. The *dead time* of a detector is defined to be the minimum length of time which must separate time incident photons in order for them to be recorded as two separate pulses. The plastic material used in these detectors is Bicron BC412, which has a rated dead time of 3.3 nanoseconds. Finally, Table 1 summarizes some miscellaneous information about the sources. Note that the source height was

Source Type	Source Strength (counts per second)	Source Height (feet)
Americium	$\sim 10^5$	5.75
Californium	$\sim 10^6$	7.0

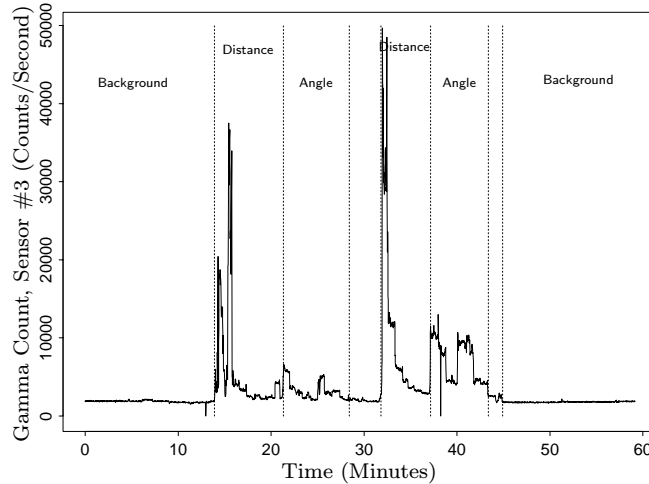
**Table 1:** Miscellaneous information about the sources used for this test.

measured from the floor to the approximate place that the source was held. Since the source was being held by a person during the experiment, there was probably a significant *unmeasured* variation in the source height over the course of the experiment.

### 3 Characterization of Radiation Sensor Antenna Pattern

The primary purpose of this experiment was to obtain data for characterizing the antenna pattern of the radiation sensors. In theory, the change in sensitivity of a radiation detector with respect to the distance  $d$  from the source to the detector is proportional to  $\frac{1}{d^2}$ . So the sensitivity of the detector decreases geometrically with increasing distance. Similarly, the change in sensitivity with respect to the angle  $\theta$  between the source and the detector is proportional to  $\cos \theta$ . Hence the sensitivity of the detector decreases with increasing angle. Since all of the measurements illustrated in Figure 3 are taken with respect to Sensor #3, only its output will be considered in this section.

The time series data gathered during this experiment for Sensor #3 is shown in Figure 5. The regions

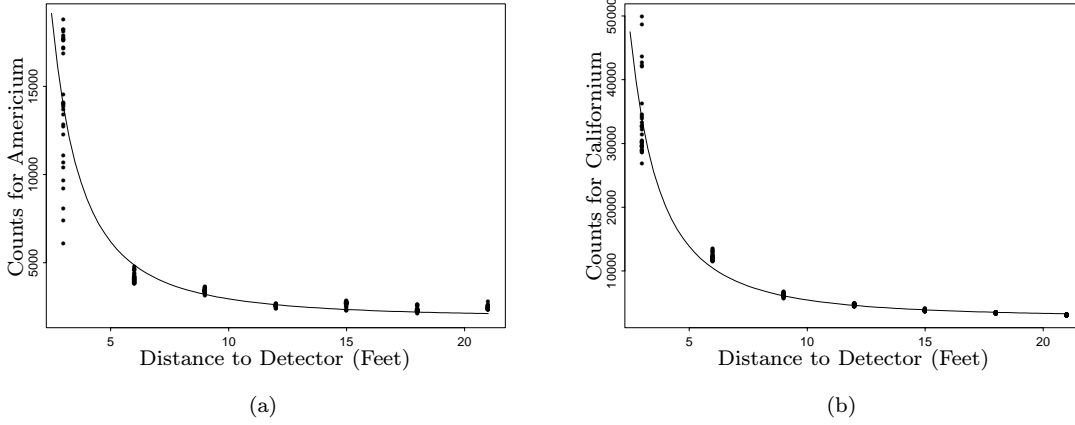


**Figure 5:** The number of gamma ray counts per second recorded by Sensor #3 over the time of the experiment.

labeled **Background** denote the times when no sources were unshielded in the room and the background radiation level was being measured. The left hand region labeled **Distance** indicates the time when

measurements 1 through 7 in Figure 3 where being made with the *Americium* source. Similarly, the right Distance region marks the same seven measurements for the *Californium* source. The left region labeled Angle shows the time when measurements 8 through 15 in Figure 3 where being taken for the *Americium* source. Likewise the right Angle region illustrates the analogous reading for the *Californium* source. Note that the large count values in these angle regions are measurements taken at 6 feet, while the small values are measurements taken at 12 feet. The small unlabeled regions depict the times when a source was either being held in the middle of the room, or being walked the length of the room.

The measurements taken in the Distance segments shown above were fitted to the model  $c = C_1 + \frac{C_2}{d^2}$ , where  $c$  is the number of counts per second, and  $d$  is the distance from the source to the detector in feet. The best fits to this model for the collected data are shown in Figure 6(a) for the *Americium* source, and in Figure 6(b) for the *Californium* source. The fitted models are of the form



**Figure 6:** (a) The number of gamma ray counts per second recorded by the detector as a function of distance from the detector for the *Americium* source. The solid circles  $\bullet$  represent the data values for each distance, and the solid line — shows the linear regression fit to those points. (b) The number of gamma ray counts per second recorded by the detector as a function of distance from the detector for the *Californium* source.

$$c = 1873.23 + \frac{107962.64}{d^2} \quad (1a)$$

$$c = 2666.05 + \frac{280268.01}{d^2}, \quad (1b)$$

where Equation (1a) is the fit to the *Americium* data, and Equation (1b) is the fit to the *Californium* data. Some statistics to assess the quality of these two fits are shown in Table 2. The Multiple  $\mathcal{R}^2$  is

	Multiple $\mathcal{R}^2$	Residual Degrees of Freedom	$\mathcal{F}$ -Statistic	$p$ -Value
Americium	0.8896	263	2120	0
Californium	0.9512	291	5669	0

**Table 2:** Statistics for assessing the quality of the fits in Equation (1) to the data shown in Figure 6.

the ratio of the variation of the fitted values to the variation of the data values, and is given by

$$\mathcal{R}^2 = \frac{\sum_{i=1}^n \left( c_1 + \frac{c_2}{d_i^2} - \frac{1}{n} \sum_{i=1}^n \left( c_1 + \frac{c_2}{d_i^2} \right) \right)^2}{\sum_{i=1}^n \left( c_i - \frac{1}{n} \sum_{i=1}^n \left( c_1 + \frac{c_2}{d_i^2} \right) \right)^2}, \quad (2)$$

where  $n$  is the total number of data points in the sample. This statistic shows how much of the variation in the data is accounted for by the model. In this case 89% of the variation in the *Americium* data is accounted for by the fit in Equation (1a), and 95% of the change in the *Californium* data is incorporated into the model in Equation (1b). The  $\mathcal{F}$ -statistic is the ratio of the covariance between the fitted values and the data values to the variation between the fitted values and the data values. In other words the statistic is given by

$$\mathcal{F} = \frac{\sum_{i=1}^n \left( \left( c_1 + \frac{c_2}{d_i^2} - \frac{1}{n} \sum_{i=1}^n \left( c_1 + \frac{c_2}{d_i^2} \right) \right) \left( c_i - \frac{1}{n} \sum_{i=1}^n \left( c_1 + \frac{c_2}{d_i^2} \right) \right) \right)}{\sum_{i=1}^n \left( c_i - \left( c_1 + \frac{c_2}{d_i^2} \right) \right)^2}. \quad (3)$$

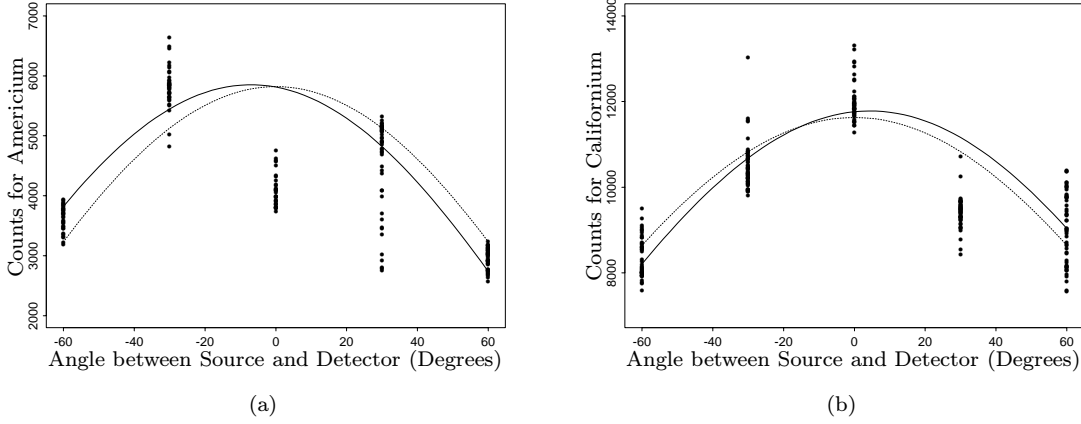
If there is a high correlation between the model values and the data values, the numerator of Equation (3) will be large, and the denominator will be small. If there is no correlation between the model and data values, then the numerator will be small and the denominator large. Therefore a large  $\mathcal{F}$ -statistic implies that the model provides a good fit to the data, while a small one means a bad fit. The underlying distribution for this statistic is the ratio of two  $\chi^2$  distributions, called a *Fisher distribution*. The Fisher distribution has two degrees of freedom associated with it, one for each of the  $\chi^2$  distributions composing it. In regression analysis one degree of freedom is the number of independent variables in the regression, which in this case is 1. The other degree of freedom is the difference between the number of data points and the number of coefficients in the model, which in this case appears in Table 2 in the column labeled **Residual Degree of Freedom**. The  $\mathcal{F}$ -statistic forms the basis for a hypothesis test assessing the quality of the fit between the model and the data. The null hypothesis is that the true slope of the regression line is 0, meaning in this instance that there is *no relationship* between number of counts and distance from the detector. The *p-Value* is the probability that the null hypothesis is true, given that the  $\mathcal{F}$ -statistic is greater than or equal to its observed value. So a small *p*-value means that there is a very small chance that the  $\mathcal{F}$ -statistic would have its observed value if there were *not* a relationship between number of counts and distance from the detector. So a small *p*-value means that the model is a good fit to the data, while a large one indicates a poor fit. Since the *p-Value* is 0 in Table 2 for both the *Americium* and *Californium* sources, the models in Equation (1) are good fits for the data.

Similarly, the data taken in the *Angle* time segments in Figure 5 were fitted to the model  $c = C_1 + C_2 \cos \theta$ , where  $c$  is the number of counts per second, and  $\theta$  is the angle between the source and the detector. Only data from the arc described by measurements at the points 9, 8, 2, 12, and 13 in Figure 3 were fitted to this model. Note that all of these points lie in a 6 foot radius of detector #3. The best linear models are of the form

$$c = 644.14 + 5175.53 \cos \theta \quad (4a)$$

$$c = 5657.46 + 5969.29 \cos \theta, \quad (4b)$$

where Equation (4a) is the fit to the *Americium* data, and Equation (4b) is the fit to the *Californium* data. These fits for the collected data are shown in Figure 7(a) for the *Americium* source, and in Figure 7(b) for the *Californium* source. Note that in Figure 7(a), the measurements taken at  $0^\circ$  were not used in determining either of the illustrated fits. Likewise in Figure 7(b), the measurements taken



**Figure 7:** (a) The number of gamma ray counts per second recorded by the detector as a function of angle between the source and detector for the *Americium* source. The solid circles  $\bullet$  represent the data values for each distance, and the dashed line --- shows the linear regression fit to those points.  
 (b) The number of gamma ray counts per second recorded by the detector as a function of distance from the detector for the *Californium* source.

at  $30^\circ$  were not used in the regression. In Figure 7(a) the solid line is a fit to the model  $\cos(\theta + 7.05)$ , and in Figure 7(b) it is the fit to  $\cos(\theta - 4.35)$ . Statistics to assess the quality of the two linear fits in Equation (4) are shown in Table 3. Based on both the Multiple  $\mathcal{R}^2$  and  $\mathcal{F}$ -Statistic, both these models are

	Multiple $\mathcal{R}^2$	Residual Degrees of Freedom	$\mathcal{F}$ -Statistic	$p$ -Value
Americium	0.6492	167	309	0
Californium	0.7783	193	677.6	0

**Table 3:** Statistics for assessing the quality of the fits in Equation (4) to the data shown in Figure 7.

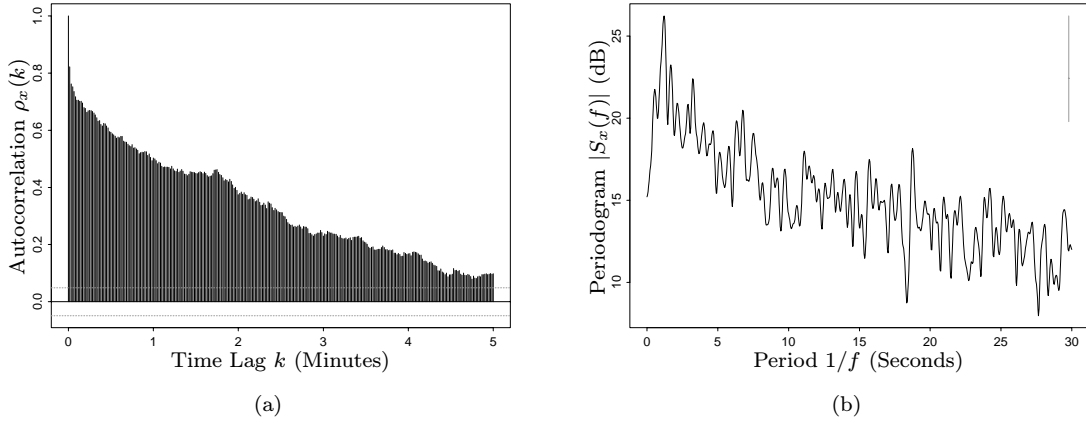
fairly good fits for the data. However, the distance models are clearly superior to the angular models.

Finally, note that raising the steel cover over the detector leads to an average increase in the sensitivity of 1.78 at 3 feet and 1.74 at 21 feet. For each distance, divide each data point taken with the cover up by a corresponding data point taken with the cover down. This leads to two sets of data, one set for each of the two distances, each set consisting of ratios of cover-up to cover-down measurements. A hypothesis test using the non-parametric Wilcoxon test was done to determine whether the means of these two data sets were equal. This means that the null hypothesis was that the two samples have the same mean. Equal means for the two samples indicates that the increase in detector sensitivity is independent of distance. The  $p$ -value of 0.375 for the Wilcoxon test indicates that at any significance less than 37%, the null hypothesis should be accepted. The level of significance is the probability of rejecting the null hypothesis when it is in fact true. So according to this test, the increase in sensitivity at the two distances is the same, with fairly high significance.

## 4 Detecting Sources with Radiation Sensors

This section discusses possible methods for detecting radioactive sources using data available from the current sensors. Classical detection theory, as studied in electrical engineering for communication, is based on the assumption that the process which generates the signal is fundamentally different from that which generates the noise. In the present application this assumption is *violated*, because both the signal and the noise are generated by the decay of unstable materials, such as uranium and plutonium, into more stable elements. This means that one must find a statistic, such as the average signal strength, whose value differs when a source is present (*i.e.*, signal) from when it is absent (*i.e.*, noise). This problem is analogous to having a communication system in which the signal is generated by heating the transmitter in order to increase its shot noise.

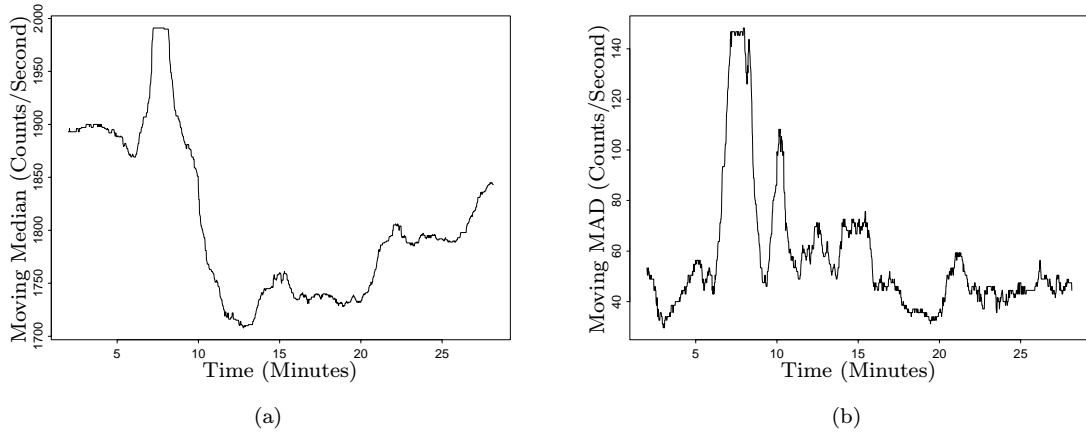
One of the simplest solutions to this detection problem is to compare the current signal strength to a reference value which represents the average signal strength when no source is present. One difficulty in the current facility is that sources are stored in the room containing the radiation sensors. Since the shielding provided by the containers is not perfect, the stored sources constitute an *unknown* contribution to the background level measured in the room. Since the number and type of sources stored is subject to change, the background level will unquestionably vary with time. Hence a source may be present in the room at all times, but it is considered part of the noise when it is in its storage container and part of the signal when it is outside its container. In order to perform the aforementioned comparison, some representation of the average background must be determined. This requires an analysis of the background data collected in the experiment discussed in Section 2. To characterize the properties of the background as a time series, two useful statistics are the *autocorrelation* and the *periodogram*. The *autocorrelation* function for a time series is found by computing the correlation coefficient  $\rho_x(k)$  for the series  $x_t$  with a time shifted version of the series  $x_{t-k}$  for different integer values of  $k$ . This function shows whether the present value of the series is independent of previous values. The *periodogram* function for a time series is found by computing the magnitude of the coefficients  $|S_x(f)|$  of the exponential Fourier series representation for the time series. This function indicates where power is distributed among the various frequency components of the time series. Strictly speaking, both of these measures assume that the time series is stationary, but in practice they seem to be good diagnostic tools even for non-stationary series. The autocorrelation and periodogram functions for the background data of the experiment are shown in Figures 8(a) and 8(b) respectively. These diagnostics reveal two features which may make the background quite difficult to model. First, the autocorrelation function reveals strong long-range serial correlations in the series. This means that consecutive time samples are *not* independent, even worse, samples separated by two minutes are not independent. This implies that the series can *not* be modeled by constructing an estimated probability distribution for the background and drawing samples as needed to compare to the current reading. By its nature, such a method assumes that each sample drawn from the distribution is independent of all other samples, which is clearly not the case here. Second, the periodogram function shows that power is rather evenly distributed over a large range of frequencies. Therefore there are no clear periodicities present in the background data. Both of these diagnostics suggest that the time series may be non-stationary. This hypothesis can be tested by looking at the moving median and moving median absolute deviation of the background. Median and median absolute deviation (MAD) are used rather than mean and standard deviation because they are far more robust in the presence of outliers. The moving median is simply the median computed over the previous  $m$  samples of the series, where  $m$  is an integer. An analogous description applies for the moving MAD. The moving median and MAD of the background for two minute windows are shown in Figures 9(a) and 9(b). These plots clearly show that both the average and the variation around the average change over time, making the background series non-stationary. The diagnostics in Figure 9 show two additional features. First, the variation of the median and the MAD over time is very large, especially in the case of the MAD where it is almost an order of magnitude. Second, the ratio of the median to the MAD squared (*cf* the mean to variance ratio) is very small, indicating a very large variance in the underlying process. In fact, these four diagnostics tend to support the hypothesis that



**Figure 8:** (a) The autocorrelation function of the background series over time lags of up to 5 minutes. (b) The periodogram function of the background series for periods of up to 30 seconds. Note that the width of the smoothing window is 0.0842385 with a 95% confidence interval of  $(-2.63618, 3.80533)$  dB. Also note that a moving series average with a 60 sample (*i.e.*, 1 minute) window size was subtracted from the background series before computing the periodogram.

the background data was generated by a non-stationary, correlated, white noise source.

In spite of the difficulties implied by the previous diagnostics, there are techniques which in principle may model the background series. Among the simplest are the autoregressive integrated moving average (ARIMA) models discussed in Box, Jenkins, and Reinsel (1994, Chapter 4). Intuitively these models allow certain types of non-stationary behavior by modeling the  $d$ th order *difference series* as a stationary autoregressive moving average (ARMA) process. The  $d$ th order *difference series* of  $x_t$  is defined  $(1 - \nabla)^d x_t$  where the operator  $\nabla^i x_t = x_{t-i}$ . Hence the second difference series is  $(1 - \nabla)^2 x_t = x_t - 2x_{t-1} + x_{t-2}$ . An ARIMA model based on the first difference is homogeneous except in level, meaning that except for vertical translation, one part of the series looks much like another part. When the second difference is used, the series is homogeneous except for level and slope. The mathematical form of an



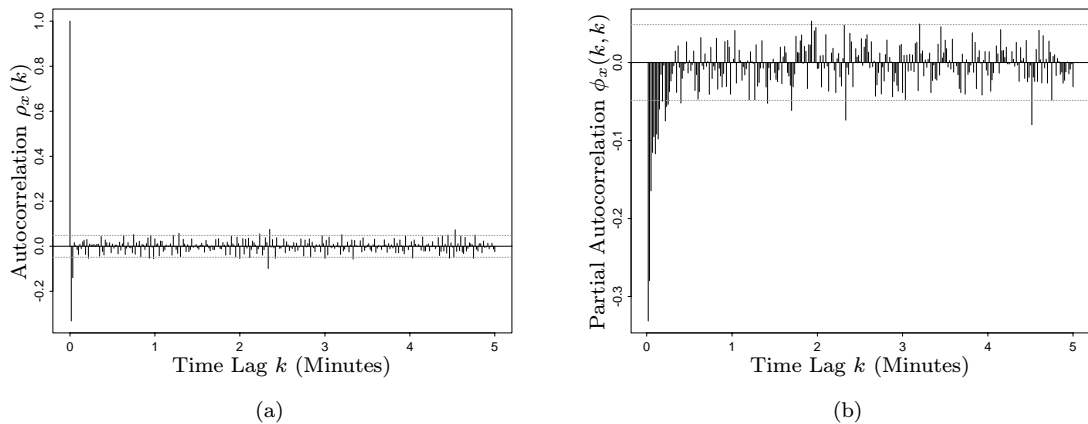
**Figure 9:** (a) The moving median of the background series for a 2 minute window. (b) The moving median absolute deviation of the background series for a 2 minute window.

ARIMA model is

$$(1 - \nabla)^d x_t - \sum_{k=1}^p \mathcal{A}_k (1 - \nabla)^d x_{t-k} = n_t - \sum_{k=1}^q \mathcal{M}_k n_{t-k}, \quad t = 1, 2, \dots, (n - d), \quad (5)$$

where  $x_i$  is the model value (*i.e.*, system output) at time  $i$ , and  $n_i$  is the value of a white noise process (*i.e.*, system input) at time  $i$ . The coefficients  $\mathcal{A}_i$  and  $\mathcal{M}_i$  weight the effect of previous outputs and inputs, respectively, on the current output. The quantity  $n$  is the number of samples in the original (*i.e.*, undifferenced) time series. When using ARIMA models there are three integers which must be chosen; the order of the difference series  $d$ ; the number of autoregressive (AR) terms  $p$ ; and the number of moving average (MA) terms  $q$ . An ARIMA model based on the  $d$ th difference series with  $p$  AR terms and  $q$  MA terms is denoted ARIMA( $p, d, q$ ). One way to choose these quantities is to plot the *autocorrelation* and *partial autocorrelation* functions for several difference series orders  $d_1, d_2, \dots, d_n$ . The proper difference order  $d_i$  can be heuristically selected as the order which yields a “small” number of “large” values in both of these plots. The *partial autocorrelation* function for a time series is found by computing the last partial autocorrelation coefficient  $\phi_x(k, k)$  for the series  $x_t$  with a time shifted version of the series  $x_{t-k}$  for different integer values of  $k$ . This function shows the correlation between  $x_t$  and  $x_{t-k}$  which is not accounted for by  $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$ . Heuristic choices for the number of AR and MA terms  $p$  and  $q$ , are the number of “large” values in the partial autocorrelation and autocorrelation functions, respectively. This heuristic is based on the analytical result that an AR process of order  $p$  has *exactly*  $p$  non-zero values in the partial autocorrelation function. Likewise, a  $q$ th order MA process has only  $q$  non-zero values in its autocorrelation function.

For the background series in question, the first and second difference series appear to have the smallest number of large values in the autocorrelation and partial autocorrelation functions. For higher order differences, the number of significant peaks seems to increase, indicating over-differencing. As an example, the autocorrelation and partial autocorrelation for the first difference of the background are shown in Figure 10. On the basis of these plots one might expect a good fit from a first difference



**Figure 10:** (a) The autocorrelation function of the first difference of the background series (*i.e.*,  $x_t - x_{t-1}$ ) over time lags of up to 5 minutes.  
 (b) The partial autocorrelation function of the first difference of the background series (*i.e.*,  $x_t - x_{t-1}$ ) over time lags of up to 5 minutes.

ARIMA model with at most four AR terms and at most two MA terms. For completeness, second difference ARIMA models were also fit to the data. There are several standard diagnostics for assessing the fit of an ARIMA model to data. Many of the diagnostics are founded on the notion of the *residual*, which is the difference between the actual data and the value of the ARIMA model, at each sample

time. For an ARIMA model, a recursive form for the residual is

$$\hat{n}_t = (1 - \nabla)^d \hat{x}_t - \left( \sum_{k=1}^p \mathcal{A}_k (1 - \nabla)^d \hat{x}_{t-k} - \sum_{k=1}^q \mathcal{M}_k \hat{n}_{t-k} \right), \quad t = 1, 2, \dots, (n - d) \quad (6)$$

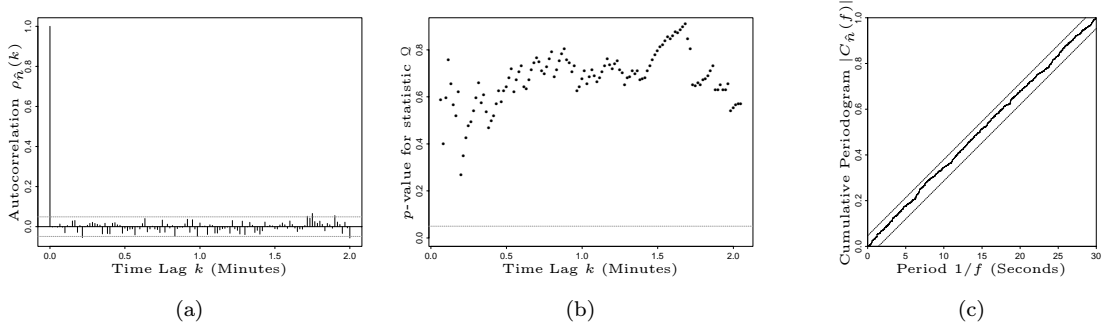
where  $\hat{x}_i$  is the actual data value at time  $i$ , and  $\hat{n}_i$  is the residual value at time  $i$ . One of the simplest diagnostics is to look at the plot of the residuals  $\hat{n}_t$  for all times  $t$ . If the model is a good fit for the data, there should be no noticeable structure in this time series, and the range of the residuals should be much smaller than the range of the data. Another straightforward diagnostic is to plot the actual data values  $\hat{x}_t$  versus the predicted values  $x_t$  for the same time period. If the model fits the data well, these plotted values should cluster around the straight line  $\hat{x}_t = x_t$ . Another test is to plot the autocorrelation of the residuals  $\rho_{\hat{n}}(k)$ . An estimator for these autocorrelations has the form

$$\rho_{\hat{n}}(k) = \frac{\sum_{t=k+1}^n \left( \hat{n}_t - \frac{1}{n} \sum_{t=1}^n \hat{n}_t \right) \left( \hat{n}_{t-k} - \frac{1}{n} \sum_{t=1}^n \hat{n}_t \right)}{\sum_{t=k+1}^n \left( \hat{n}_t - \frac{1}{n} \sum_{t=1}^n \hat{n}_t \right)^2}, \quad k = 0, 1, 2, \dots, k \leq (n - d), \quad (7)$$

where  $k$  is the time lag computed. For a good model fit the time series of residuals should be a white noise process, hence the correlation of residuals from different time lags  $k$  should be very small. Rather than considering the autocorrelations individually, another useful indicator is the first  $k$  autocorrelations taken together. One statistic of this type, which is used to judge model adequacy, is the sum of the squares of the first  $k$  autocorrelations

$$\mathcal{Q} = (n - d) \sum_{k=1}^k \rho_{\hat{n}}^2(k). \quad (8)$$

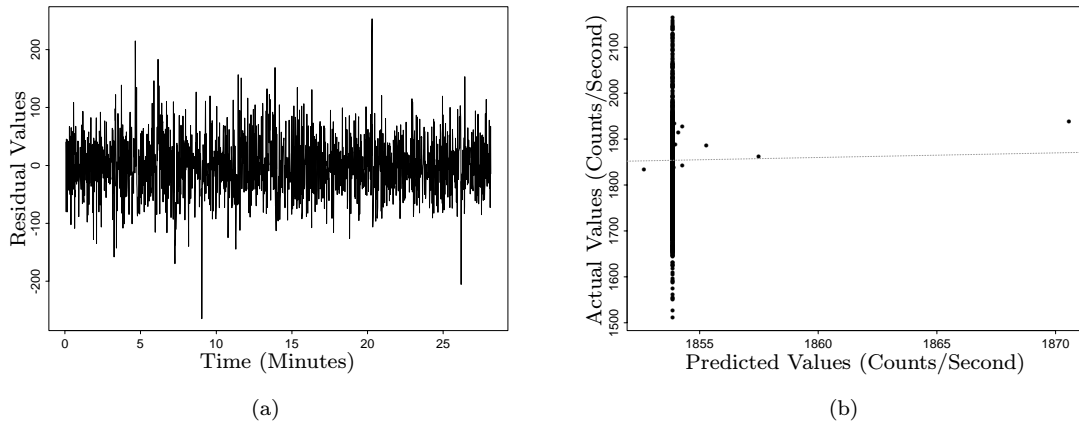
The underlying distribution for  $\mathcal{Q}$  is a  $\chi^2$  distribution with  $(k - p - q)$  degrees of freedom. This statistic is used in a “portmanteau” test of the hypothesis that the model is a good description of the data. For every desired time lag  $k$ , a  $p$ -value is returned for this hypothesis test. The  $p$ -value specifies the *maximum* significance at which the null hypothesis should be accepted. So if the desired level of significance is less than the  $p$ -value, accept the null hypothesis, otherwise reject it. The level of significance is the probability of rejecting the null hypothesis when it is true. The relevant diagnostic resulting from this hypothesis test is a plot of the  $p$ -values associated with  $\mathcal{Q}$  for a number of different time lags  $k$ . If the model fits the data well, these  $p$ -values should be fairly close to 1, indicating that the null hypothesis (*i.e.*, the model fits the data well) should be accepted for any reasonable significance. In a sense the  $p$ -value is the probability that the current value of  $\mathcal{Q}$  implies that the model fits the data. A final measure of fit quality is the *cumulative periodogram*  $|C_{\hat{n}}(f)|$  of the residuals. This is computed for frequency  $f_i$  by summing the normalized magnitudes of the exponential Fourier series coefficients  $|S_{\hat{n}}(f_i)|$  for all frequencies up to  $f_i$ . This function reveals any periodic nonrandomness in the residuals. When the model fit is good, all the values of  $|C_{\hat{n}}(f)|$  are scattered around the straight line joining the points (0,0) and (0.5,1). Of the models tried, a first difference model with three AR terms and one MA term (*i.e.*, an ARIMA(3,1,1) model) gave the best fit based on the three standard diagnostics shown in Figure 11. From the above discussion about these diagnostics, ideally all the autocorrelations, except the first, should lie inside the dotted lines, all the  $p$ -values should be close to 1, and the cumulative periodogram should always lie between the two dotted lines. These model diagnostics are discussed in detail in Box *et al.* (1994, Chapter 8). In spite of the good performance of the ARIMA(3,1,1) model with respect to these diagnostic statistics, it actually provides a very poor fit to the experimental data, which is clearly shown in Figure 12. Figure 12(a) graphs the residuals for the ARIMA(3,1,1) model at each sample time. This plot reveals that the range of the residuals is the same as the range of the first difference series of the background, which is a clear indication that the model is not capturing the character of



**Figure 11:** (a) The autocorrelation function of the residuals for the background data when fit to an ARIMA(3,1,1) model.  
 (b) The probability that the ARIMA(3,1,1) model would generate the observed autocorrelations up to lag  $k$  if it were a good fit to the data. This is a *portmanteau test* of the hypothesis that the model is adequate.  
 (c) The cumulative of the residuals for the data when fit to an ARIMA(3,1,1) model. The cumulative periodogram at frequency  $f_i$  is the sum of the spectral densities at all frequencies up to  $f_i$ .

the data. Figure 12(b) plots the predictions of the ARIMA(3,1,1) model versus the actual background data. Ideally, the points in this plot should cluster around the dotted line running through the center. Instead this graph indicates that the range of the predictions is much smaller than the range of the actual data, and in fact most of the predicted values lie in a very narrow window around the value 1854. These plots indicate that the ARIMA(3,1,1) model captures the mean of the series but fits *none* of its variance. Both of these plots make it clear that the ARIMA(3,1,1) model is a *very poor* model of the actual background series.

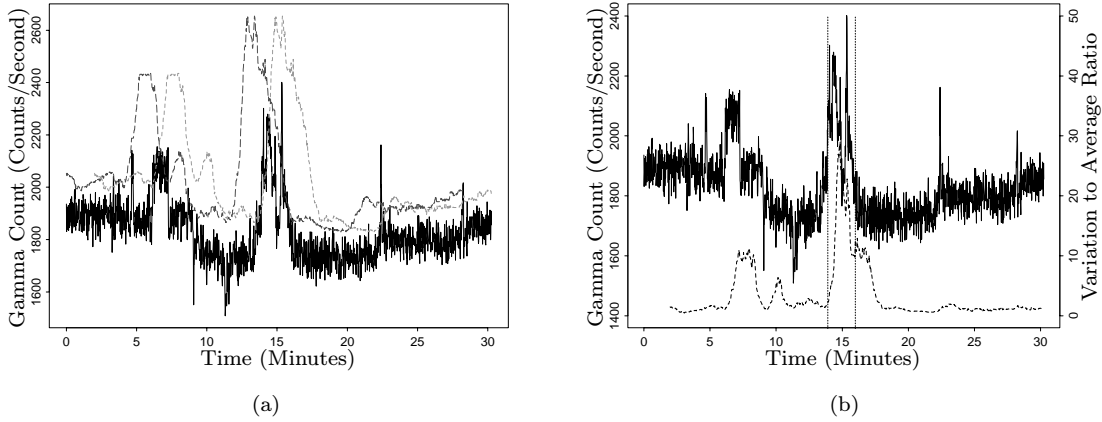
Since it does not appear that the entire background series can be well modeled by any reasonably simple model, the next logical step is to try to model some smaller window of the data. Conceptually this model or summary can be compared to the current data value, and a decision concerning the presence of a source made on the basis of this comparison. Intuitively, one of the simplest versions of this approach



**Figure 12:** (a) The residuals for the background data when fit to an ARIMA(3,1,1) model.  
 (b) The background values predicted by the ARIMA(3,1,1) model versus the actual background data.

involves setting a threshold based on some “window” of the data, and deciding that a source has been detected when the actual sensor signal exceeds the computed threshold signal. Since both the median (*cf* mean) and the median absolute deviation (*cf* standard deviation) of the background signal are time-varying, as illustrated in Figure 9; and since the signal deviation is large, it makes sense to set a threshold equal to the median plus some number of median absolute deviations. A slightly different, but still simple, approach is to look for some characteristic change in the model or summary. Methods of this type are discussed in Basseville and Nikiforov (1993). A straightforward method of this type is to decide that a source has been detected when the ratio of the median absolute deviation squared (*cf* variance) to the median abruptly changes in value.

A test series was constructed by inserting the measurements taken while standing in the middle of the high-bay with the *Americium* source and while walking the length of the high-bay with that source, between the two sets of background measurements. This test series is a worst-case scenario in the sense that the distance from the source to all detectors is as large as possible, and the strength of the source is as small as is likely. The results for both strategies outlined above are shown in Figure 13. The first approach uses a moving window of 2 minutes worth of data. For this method, the previous 2



**Figure 13:** (a) The test series plotted with both the on-set and off-set detection thresholds.  
(b) The test series plotted with the ratio of variation to average threshold.

minutes of data are used to compute a threshold for detecting the arrival of a source in the room. This threshold was constructed by taking the median value of the previous 2 minutes of data and adding 3 times the median absolute deviation for the previous 2 minutes of data. When the actual sensor data goes above this threshold then a source is considered to be present. In order to detect the departure of a source, a similar threshold is constructed using the following 2 minutes of data. When the sensor data drops below the second threshold the source is considered to be absent. Since the two thresholds are identical except for a time shift, they can both be computed from the previous 2 minutes of data. However, the first threshold is computed at the *current* time, while the second is computed at a time *2 minutes previous*. This method is illustrated in Figure 13(a), and it looks fairly effective at catching when a source is present. Note the false alarm at about 6 minutes, but this change appears to the naked eye like a source coming into the room. The second approach also uses a moving window composed of 2 minutes of data. For this method, the previous 2 minutes of data is used to compute the ratio of the median absolute deviation squared to the median 2 minutes ago. When the value of this statistic abruptly increases, a source has entered the room, and when it decreases one has left the facility. The resulting detection threshold is graphed in Figure 13(b). Note that threshold lags the actual signal by 2 minutes since past data is used to compute the present ratio. This method also appears quite effective except for the false alarm at 6 minutes. Note that the pair of vertical dotted lines denotes the region in

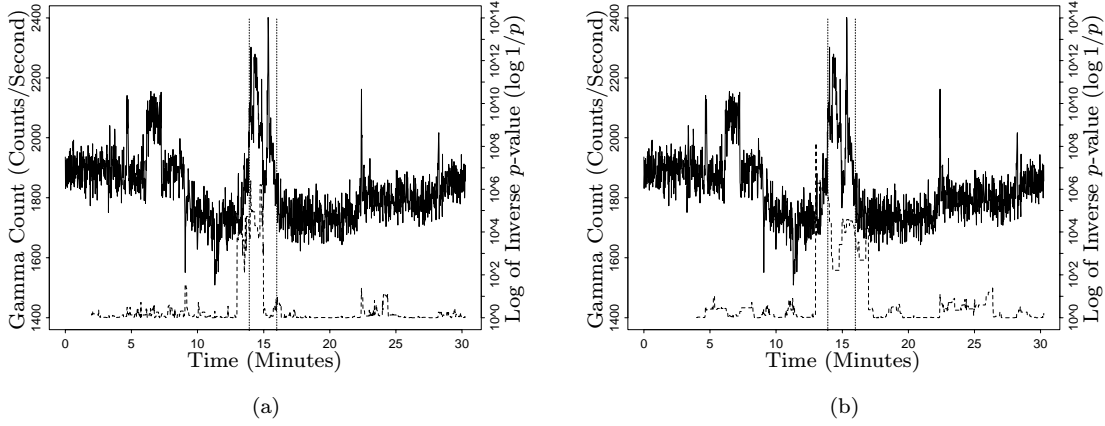
which a source is present in the signal.

A slightly different approach to detection is to split each window in half, compute some statistic in each half of the window, and compare the similarity of these two statistics. If the two statistics are different then a source is present. Note that each window half will be called a sample, and the elements in each window half will be called observations. Denote the sample from one half of the window by the vector  $\mathbf{x}$  drawn from the random variable  $\mathcal{X}$ , and the sample from the other half by  $\mathbf{y}$  drawn from  $\mathcal{Y}$ . A very well known test for comparing cumulative distribution functions is the *Kolmogorov-Smirnov test*, which is discussed in Conover (1980, Chapter 6) and Hollander and Wolfe (1973, Chapter 10). The cumulative distribution function (cdf) is a mapping which assigns the probability that a random variable  $\mathcal{Z}$  is less than or equal to some specific value  $z$  for all possible  $z$  values, in other words  $F_{\mathcal{Z}}(z) = \mathcal{P}\{\mathcal{Z} \leq z\}$ . The test statistic is

$$\mathcal{K} = \sup_{-\infty < z < \infty} |F_{\mathcal{X}}(z) - F_{\mathcal{Y}}(z)| \quad (9)$$

where  $F_{\mathcal{X}}(z)$  is the cdf computed from the sample  $\mathbf{x}$  (*i.e.*, the first window half), and  $F_{\mathcal{Y}}(z)$  is the cdf calculated for sample  $\mathbf{y}$  (*i.e.*, the second window half). In both halves of the window the cdf is approximated by computing the histogram in that half and computing the value of the cdf for the  $i$ th bin by taking the cumulative sum of the number of elements in each bin of the histogram up to the  $i$ th bin. Intuitively this  $\mathcal{K}$  statistic is the maximum difference between the two distributions. The underlying probability density function for  $\mathcal{K}$  is  $1 - 2 \sum_{l=1}^{\infty} (-1)^{l-1} e^{-2l^2 w^2}$  for all  $w > 0$ . The null hypothesis is  $F_{\mathcal{X}}(z) = F_{\mathcal{Y}}(z)$  for all  $z$ , and the alternative is  $F_{\mathcal{X}}(z) \neq F_{\mathcal{Y}}(z)$  for at least one value of  $z$ . The null hypothesis is *accepted* if the test statistic  $\mathcal{K}$  is less than some threshold  $\mathcal{T}_{\mathcal{K}}$  which is chosen based on the desired false alarm rate  $\alpha$ . Recall that a hypothesis test returns a  $p$ -value which specifies the *maximum* significance at which the null hypothesis should be accepted, and that the level of significance is the probability of accepting the alternative hypothesis when the null hypothesis is true. This means that selecting a significance level in hypothesis testing is equivalent to choosing a maximum false alarm rate in detection. If the desired false alarm rate is less than the  $p$ -value, accept the null hypothesis, otherwise reject it. Hence a  $p$ -value close to one means the two cdfs are equal (*i.e.*, no source present), and a  $p$ -value near zero indicates they are not equal (*i.e.*, source present). Note that in order for this test to return a small  $p$ -value the *entire time* a source is present, the cdf when a source is absent must differ from the cdf when a source is present, and the cdf must change fairly quickly *while* the source is present. The results for this method are shown in Figure 14. The right hand axis in both these plots is the logarithm of the inverse of the  $p$ -value, rather than the  $p$ -value itself. This means that high values indicate the presence of a source and low values the absence of one. Figure 14(a) shows this statistic for a 2 minute window size, and Figure 14(b) plots it for a 4 minute window. Note that the threshold *should* lag the actual signal since past data is used to compute the present  $p$ -value. The fact that the threshold leads the signal in both plots indicates a false alarm in the portion of the background that directly precedes the appearance of the source. Since this test signal is pieced together, this false alarm may be due to an error in deciding when the background signal ended. In any case it seems reasonable to state that both window sizes catch the leading edge of the source signal. Clearly the 2 minute window does not catch the trailing edge of the source signal, while the 4 minute window appears to catch it.

Many similar detection strategies are discussed in Gibson and Melsa (1975). One of these strategies is called the *Spearman Rho Detector*. This detector is based on the Spearman Rho test which is discussed in Conover (1980, Chapter 5) and Hollander and Wolfe (1973, Chapter 8). In the context of the current problem this detector computes a correlation coefficient between the two window halves and then tests to determine whether these two samples are correlated or not. Intuitively one would think that in the present situation the correlation between window halves would be non-zero with no source and zero with a source. Since the autocorrelation estimator in Equation (7) depends on the probability density



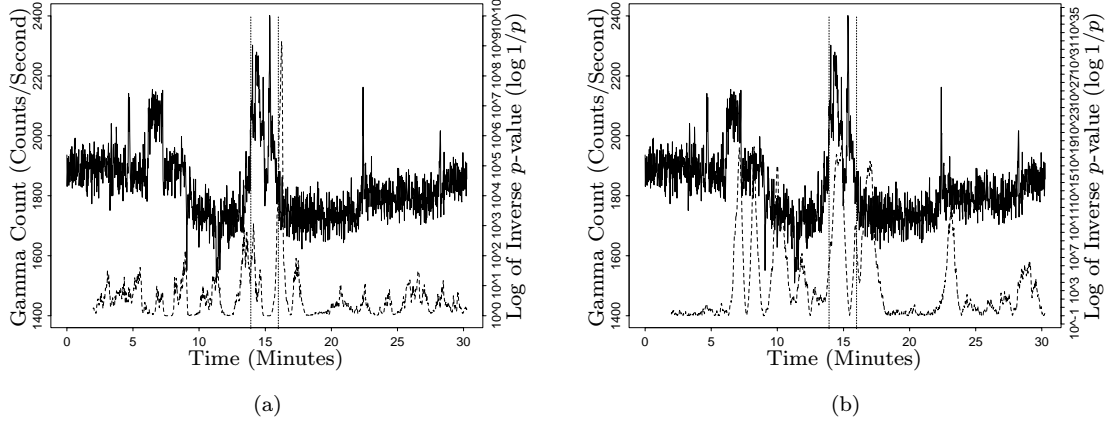
**Figure 14:** (a) The test series plotted with the Kolmogorov-Smirnov threshold for a 120 second window. Note that in this plot *high* values indicate the *presence* of a source.  
 (b) The test series plotted with the Kolmogorov-Smirnov threshold for a 240 second window. Note that in this plot *high* values indicate the *presence* of a source.

from which the samples are drawn, Spearman developed a different correlation measure given by

$$\mathcal{R} = \frac{\sum_{i=1}^n \left( R_{\mathbf{x}}(x_i) - \frac{1}{2}(n+1) \right) \left( R_{\mathbf{y}}(y_i) - \frac{1}{2}(n+1) \right)}{\frac{1}{12}n(n+1)(n-1)}, \quad (10)$$

which is independent of the underlying distribution. In this equation  $R_{\mathbf{x}}(x_i)$  and  $R_{\mathbf{y}}(y_i)$  are the *ranks* of the observations in the two samples  $\mathbf{x}$  and  $\mathbf{y}$ ,  $x_i$  and  $y_i$  respectively for  $i = 1, \dots, n$ . Specifically,  $R_{\mathbf{x}}(x_i) = 1$  if  $x_i$  is the smallest observation in magnitude,  $R_{\mathbf{x}}(x_i) = 2$  if  $x_i$  is the second smallest in magnitude and so on. The observation with the largest magnitude has rank  $n$ . The exact probability density of  $\mathcal{R}$  can not be analytically determined, but it has been shown that asymptotically the random variable  $\mathcal{Z} = \sqrt{n-1} \mathcal{R}$  has a Gaussian probability density function. The null hypothesis is that the samples  $\mathbf{x}$  and  $\mathbf{y}$  are independent, and the alternative is that the standard autocorrelation coefficient  $\rho$  is greater than zero. Note that in general, independence implies that  $\rho = 0$ , but that the reverse is *false*. The hypothesis test returns a *p*-value which if near one means that the window halves are correlated (*i.e.*, no source present), and if near zero means that the two window halves are uncorrelated (*i.e.*, source present). The result of using this detector is shown in Figure 15(a). The Spearman detector seems to trigger on the leading and trailing edges of the time interval containing the source. For some reason the leading edge spike is not markedly higher than some of the spikes in the background intervals. Note that the presence of spike trains in the background intervals indicates that the window halves are probably not independent even when a source is absent.

Another strategy discussed in Gibson and Melsa (1975) is the *Wilcoxon Rank Sum Detector*. This detector is based on the Wilcoxon Rank Sum test which is discussed in Conover (1980, Chapter 5) and Hollander and Wolfe (1973, Chapter 4). Conceptually this detector determines whether the cdfs in the two window halves are equal by finding the probability that the sample from one half is greater than the sample from the other half. If this probability is equal to  $\frac{1}{2}$  then the cdfs are the same, if not they are different. Note that this implicitly assumes that the two samples  $\mathbf{x}$  and  $\mathbf{y}$  are independent. To perform this test, combine all of the observations from the two samples, and rank the  $2n$  observations in this combined sample. As described above, ranking amounts to sorting the observations in ascending order by magnitude and replacing the smallest observation by the number 1, the second smallest by the number



**Figure 15:** (a) The test series plotted with the Spearman threshold for a 120 second window. Note that in this plot *high* values indicate the *presence* of a source.  
 (b) The test series plotted with the Wilcoxon threshold for a 120 second window. Note that in this plot *high* values indicate the *presence* of a source.

2, and so on up to the largest which is replaced by  $2n$ . The notation  $\mathbf{x} \oplus \mathbf{y}$  denotes the concatenation of the two samples  $\mathbf{x}$  and  $\mathbf{y}$ . The test statistic  $W$  is the sum of the  $n$  ranks from *one* of the two original samples, in other words  $W_{\mathbf{x}}$  is the sum of the ranks of the  $n$   $x_i$  values  $W_{\mathbf{x}} = \sum_{i=1}^n R_{\mathbf{x} \oplus \mathbf{y}}(x_i)$ . The exact probability density of  $W$  can not be analytically determined, but it has been shown that asymptotically it has a Gaussian probability density function. The null hypothesis is that the probability that the observations from one sample are greater than the other  $\mathcal{P}\{X < Y\}$  is equal to  $\frac{1}{2}$ , and the alternative is that this probability is not  $\frac{1}{2}$ . The hypothesis test returns a  $p$ -value which if near one means the two cdfs are equal (*i.e.*, no source present), and if near zero indicates they are not equal (*i.e.*, source present). The result of using this detector is shown in Figure 15(b). The Wilcoxon detector also triggers on the leading and trailing edges of the source interval. However, it also triggers strongly on several places in the background, among them the anomalous change in the background at about 6 minutes. Some of the problems with the Wilcoxon detector may result from the fact that many pairs of window halves are *not* independent as indicated by Figure 15(a). Finally, note that all three of these detection strategies assume that the observations *in a particular window half* are independent. Among other things this means that there is no serial correlation in any specific sample. For the current data set this assumption is *violated* in many window halves. One way to eliminate the serial correlation is to use the first difference series instead of the actual time series. Unfortunately, *all* of the above detection methods perform *much poorly* with the difference series than they do with the undifferenced series. We believe that part of the explanation for this observation is the fact that the difference series contains no information about changes in the mean, and the loss of this information makes the detection problem much harder.

## 5 Locating Sources with Radiation Sensors

This section discusses possible methods for locating radioactive sources using data available from the current sensors. All of the methods considered require a model relating the source location to the detected count rate. One possible model is considered in the first subsection. The problem of source location can be viewed as the search for a mapping from detected count rate to source location. Therefore given a mapping from source location to detected count rate, the problem reduces to inverting this mapping. The solution to this problem will be framed in two ways; first as a root finding problem; and

second as a nonlinear least squares estimation problem. Note that the first formulation constitutes a special case of the second.

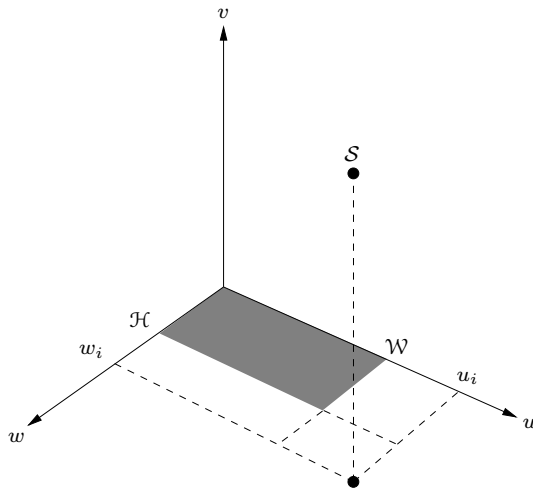
### 5.1 Sensor Model

In this section we discuss the model that relates the detected count rate from a point source to the position of that source based on the analysis in Tsoulfanidis (1983, Chapter 8). For the  $i$ th detector, the relationship between the detected count rate  $\mathcal{D}_i$  and the source position  $(u_i, v_i, w_i)$  relative to that detector is given by

$$\begin{aligned} \mathcal{D}_i &= \frac{\Omega_i(u_i, v_i, w_i) \mathcal{S} \epsilon_i \mathcal{F}_i + \mathcal{B}_i}{1 + \tau_i \Omega_i(u_i, v_i, w_i) \mathcal{S} \epsilon_i \mathcal{F}_i} \\ \Rightarrow \mathcal{D}_i &= \mathcal{M}_i(u_i, v_i, w_i) \end{aligned} \quad (11)$$

for  $i = 1, 2, \dots, m$ , where in our case  $m = 4$ . In this equation  $\mathcal{S}$  is the actual source strength,  $\epsilon_i$  is the detector efficiency,  $\tau_i$  is the dead time,  $\mathcal{F}_i$  is the product of all the correction factors (*e.g.*, absorption and backscattering),  $\mathcal{D}_i$  is the total number of counts per unit time actually detected,  $\mathcal{B}_i$  is the number of counts per unit time which constitute the background, and the view factor  $\Omega_i(\cdot)$  is the ratio of the number of particles which actually enter the detector to the total number of particles emitted by the source. The quantity  $\Omega_i(\cdot)$  can also be thought of as the solid angle subtended by the detector for a particular source location, which ranges between 0 and  $4\pi$ . For these detectors, the efficiency is rated at  $\sim 10\%$  for a  $\text{Cs}^{137}$  source. The *dead time* of a detector is defined to be the minimum length of time which must separate incident photons in order for them to be recorded as two separate pulses. The photoluminescent material used in these detectors has a dead time of 3.3 nanoseconds. Note that there is one equation of this form for each of the  $m$  detectors. We will use the right hand side of Equation (11) as an approximation for the count rate given a particular position  $(u_i, v_i, w_i)$ , and we denote this function by  $\mathcal{M}_i(\cdot)$ .

For a point source and a rectangular detector of finite size, the most general relative position is shown in Figure 16. The solid angle  $\Omega_i(\cdot)$  subtended by a detector of width  $\mathcal{W}$  and height  $\mathcal{H}$  for a point

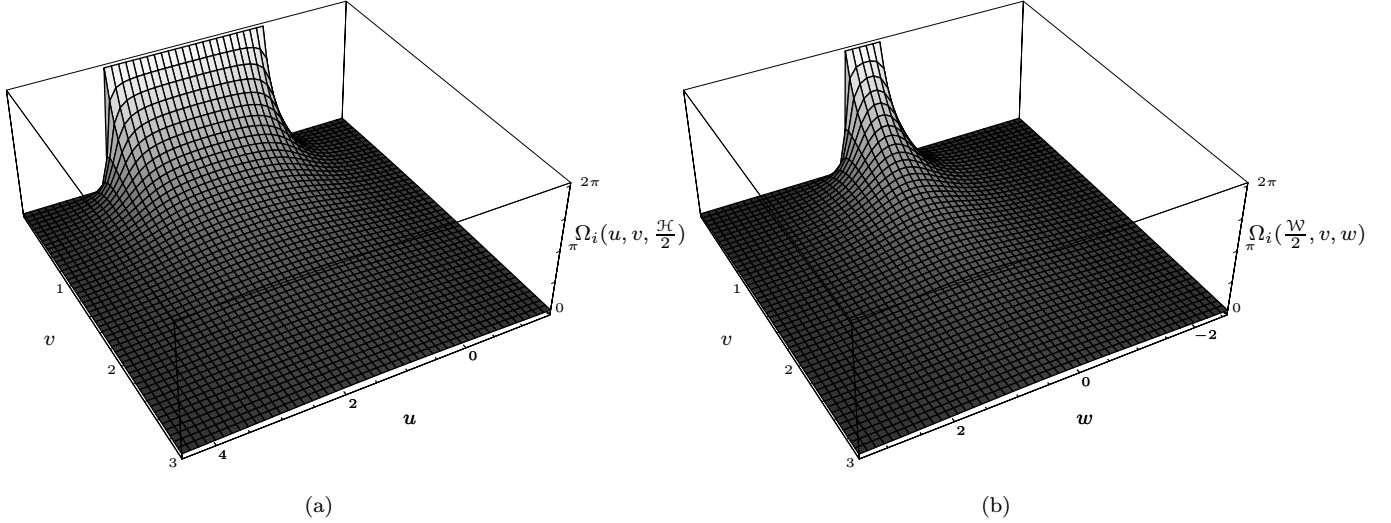


**Figure 16:** The relative position between a point source at position  $S$  and a rectangular detector of width  $\mathcal{W}$  and height  $\mathcal{H}$ . The projections of the source location onto the  $u$  and  $w$  axes are  $u_i$  and  $w_i$  respectively.

source located at  $(u_i, v_i, w_i)$ , relative to the  $i$ th detector, is given by

$$\begin{aligned} \Omega_i(u_i, v_i, w_i) = & \arctan\left(\frac{u_i w_i}{|v_i| \sqrt{u_i^2 + w_i^2 + v_i^2}}\right) - \arctan\left(\frac{(u_i - \mathcal{W}) w_i}{|v_i| \sqrt{(u_i - \mathcal{W})^2 + w_i^2 + v_i^2}}\right) \\ & - \arctan\left(\frac{u_i (w_i - \mathcal{H})}{|v_i| \sqrt{u_i^2 + (w_i - \mathcal{H})^2 + v_i^2}}\right) + \arctan\left(\frac{(u_i - \mathcal{W}) (w_i - \mathcal{H})}{|v_i| \sqrt{(u_i - \mathcal{W})^2 + (w_i - \mathcal{H})^2 + v_i^2}}\right) \end{aligned} \quad (12)$$

as shown in Gotoh and Yagi (1971). The detectors that were used have a height  $\mathcal{H}$  of 0.83 feet and a width  $\mathcal{W}$  of 3.00 feet. The minimum value of  $\Omega_i(u, v, w)$  is 0, and its maximum value is  $2\pi$  since even when the source is touching a planar detector, only half of the emitted photons will actually strike the detector. The variation of the solid angle  $\Omega_i(u, v, w)$  with position is illustrated in Figure 17. Figure 17(a) shows the variation in the  $u$ - $v$  plane with the source fixed along the  $w$ -axis at  $\frac{\mathcal{H}}{2}$ , and



**Figure 17:** The variation of the solid angle  $\Omega_i(u, v, w)$  in the (a)  $u$ - $v$  plane (b)  $w$ - $v$  plane. The third dimension was fixed at (a)  $w = \frac{\mathcal{H}}{2}$  (b)  $u = \frac{\mathcal{W}}{2}$ .

Figure 17(b) plots the change in the  $w$ - $v$  plane with the source fixed along  $u$  at  $\frac{\mathcal{W}}{2}$ . Note that the solid angle is always positive and approaches  $2\pi$  as the source gets closer to the detector. Also the solid angle decreases monotonically both as the source-detector distance increases and as the source-detector angle increases. We define source-detector angle as the angle measured relative to the line perpendicular to the center of the detector. Hence this angle is 0 when the source is directly above the center of the detector.

It is clear from Figure 16 that the coordinate system used to derive Equation (12) is detector centered, **not** room centered. Since the desired answer is the source location in room centered coordinates, a transformation must be made between these two coordinate systems. The required transformation converts room centered coordinates  $(x, y, z)$  into detector centered coordinates  $(u_i, v_i, w_i)$ , which allows Equation (12) to be properly evaluated. One way to construct this transformation is by moving each of the detectors from their actual location to the selected origin of the room. This is accomplished by first rotating the sensor about the  $w$ -axis in Figure 16, and then translating the rotated sensor to the origin

of the room. This transformation can be written in matrix form as

$$\begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix} = \begin{pmatrix} \cos \mathcal{A}_i & \sin \mathcal{A}_i & 0 & \mathcal{T}_{x,i} \cos \mathcal{A}_i + \mathcal{T}_{y,i} \sin \mathcal{A}_i \\ -\sin \mathcal{A}_i & \cos \mathcal{A}_i & 0 & \mathcal{T}_{y,i} \cos \mathcal{A}_i - \mathcal{T}_{x,i} \sin \mathcal{A}_i \\ 0 & 0 & 1 & \mathcal{T}_{z,i} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \quad (13)$$

where  $\mathcal{T}_{x,i}$ ,  $\mathcal{T}_{y,i}$ , and  $\mathcal{T}_{z,i}$  are the translations along the room centered  $x$ ,  $y$ , and  $z$  directions, respectively, for the  $i$ th detector. The quantity  $\mathcal{A}_i$  is the rotation angle around the  $w$ -axis for the  $i$ th detector. Note that the source location in room centered coordinates is  $(x, y, z)$ . Also note that the source strength  $\mathcal{S}$  is independent of position, so it is unchanged under coordinate transformation.

## 5.2 Locating Sources using Root Finding

This subsection formulates the source location problem as a root finding problem using the sensor model. Recall from Figure 1 that there are four radiation sensors in the current facility, hence there are four functions in the form of Equation (11). Rewrite each of these equations in the form  $f_i(x, y, z, \mathcal{S}) = 0$  for  $i = 1, \dots, 4$ . All four of these equations taken together yield a vector equation of the form  $\mathbf{f}(x, y, z, \mathcal{S}) = \mathbf{0}$ . The only valid location(s)  $(x, y, z)$  and source strength(s)  $\mathcal{S}$  are those which cause  $\mathbf{f}(\cdot)$  to be close to zero. So intuitively, locating the source is a multidimensional root finding problem in which the roots are points specifying a location and a source strength. The function  $\mathbf{f}(\cdot)$  is called the *residual* and the *states* are collectively denoted by the vector  $\mathbf{e}^\dagger = [x \ y \ z \ \mathcal{S}]$ . The root finding problem consists of finding a state  $\mathbf{e}$  which minimizes the residual  $\mathbf{f}(\mathbf{e})$ . One way to solve this problem is to compute the second order Taylor series expansion of  $\mathbf{f}(\mathbf{e})$  about the point  $\mathbf{e}$

$$f_i(\mathbf{e} + \delta \mathbf{e}) = f_i(\mathbf{e}) + \sum_{j=1}^m \frac{\partial f_i}{\partial e_j} \delta e_j + O(\delta \mathbf{e}^2), \quad i = 1, 2, \dots, m. \quad (14)$$

Neglecting terms of order  $\delta \mathbf{e}^2$  and higher and setting  $\mathbf{f}(\mathbf{e} + \delta \mathbf{e}) = \mathbf{0}$ , we obtain a set of linear equations for the corrections  $\delta \mathbf{e}$  that move each residual  $f_i(\mathbf{e})$  toward zero simultaneously. For the  $k$ th iteration of the algorithm, the vector form of these equations is

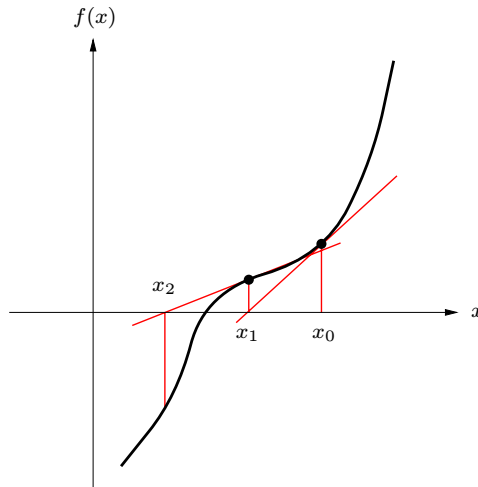
$$\delta \mathbf{e}_k = -\mathbf{J}_{\mathbf{f}}^{-1}(\mathbf{e}_k) \mathbf{f}(\mathbf{e}_k), \quad (15)$$

where  $\mathbf{J}_{\mathbf{f}}(\mathbf{e}_k)$  is the Jacobian matrix of the system  $\frac{\partial \mathbf{f}(\mathbf{e}_k)}{\partial \mathbf{e}_k}$ . The corrections are added to the solution vector giving the update rule

$$\mathbf{e}_{k+1} = \mathbf{e}_k + \alpha_k \delta \mathbf{e}_k, \quad (16)$$

where  $\alpha_k \in (0, 1]$  is a weighting factor to keep the algorithm from overshooting the solution. This algorithm for root solving is commonly known as the *Newton-Raphson method*, it is a special case of the *Gauss-Newton algorithm*, and is discussed in Fletcher (1987, Chapter 6). Intuitively, the Newton-Raphson method fits a tangent hyperplane to each component  $f_i(\cdot)$  at the point  $\mathbf{e}_k$  for each equation  $i = 1, \dots, m$  and then seeks the intersection of these  $m$  hyperplanes. The actual root is approximated by this intersection point, which then becomes the next point  $\mathbf{e}_{k+1}$  at which the hyperplanes are constructed. In one dimension the iterative process is illustrated by Figure 18. The process begins at  $x_0$  by constructing the tangent line to  $f(x)$  at this point. The intersection of this tangent with the  $x$ -axis is found and denoted  $x_1$ . The value of  $f(x)$  is computed at this new point and then the tangent to  $f(x)$  at  $x_1$  is found. The algorithm proceeds in this manner until the root is located.

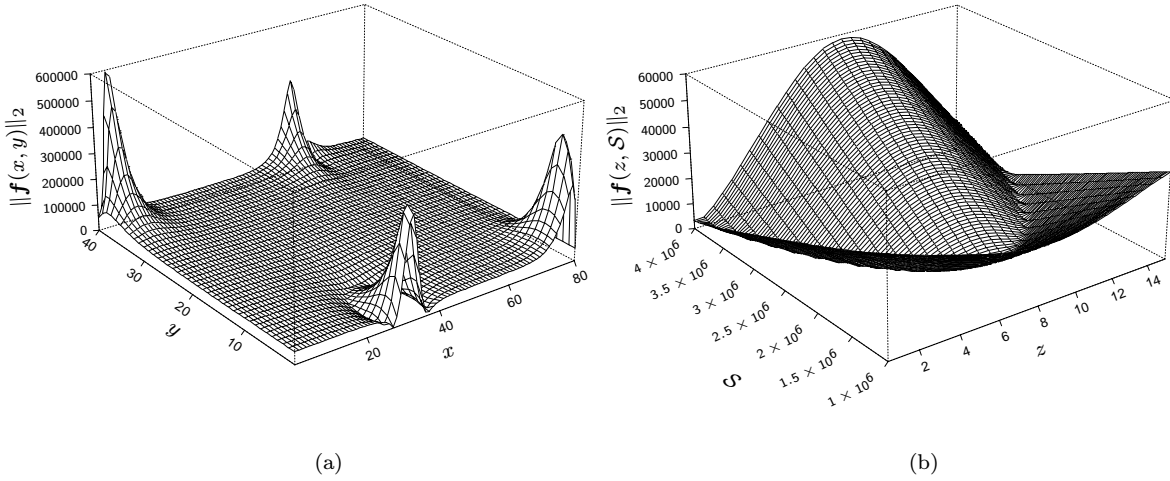
One reasonable way to choose  $\alpha_k$  is outlined in Press *et al.* (1992, Chapter 9). The strategy consists of choosing a value of  $\alpha_k$  at each iteration such that the function  $r = \frac{1}{2} \mathbf{f}^\dagger \mathbf{f}$  is reduced. Note that every



**Figure 18:** The progression of iterations in the Newton-Raphson algorithm in one dimension.

root of  $\mathbf{f}(\mathbf{e}) = \mathbf{0}$  is also a local minimum of  $r$ , but there may be local minima of  $r$  which are *not* roots of  $\mathbf{f}(\mathbf{e}) = \mathbf{0}$ . This means that trying to find the roots of  $\mathbf{f}(\mathbf{e}) = \mathbf{0}$  by simply minimizing  $r$  is likely to perform very poorly. Also note that the Newton step direction defined by  $\delta \mathbf{e}$  is a descent direction of  $r$  since  $\nabla r^\dagger \delta \mathbf{e} = (\mathbf{J}_f^\dagger \mathbf{f})^\dagger (-\mathbf{J}_f^{-1} \mathbf{f}) = -\mathbf{f}^\dagger \mathbf{f} < 0$ . The algorithm begins by trying the full Newton step (*i.e.*,  $\alpha_k = 1$ ) in the direction defined by  $\delta \mathbf{e}_k$  and checking to see whether this step reduces  $r$ . If  $r$  is reduced then this step is taken, if *not* then the algorithm backtracks along the direction specified by  $\delta \mathbf{e}_k$  until a value of  $\alpha_k$  is found which does reduce  $r$ . Since the Newton step direction  $\delta \mathbf{e}_k$  is a descent direction of  $r$ , backtracking along  $\delta \mathbf{e}_k$  is guaranteed to find an acceptable step size  $\alpha_k$ . This modification of Newton's method converges to a root from almost any starting point. This is in sharp contrast to pure Newton's method which tends to diverge to infinity unless it is started very close to a root. Other root finding methods which have almost global convergence properties are discussed in Fletcher (1987, Chapter 6), Press *et al.* (1992, Chapter 9) and Dennis and Schnabel (1983, Chapter 6).

Of course there are problems that any root finding method will have difficulty with. Some of the principle difficulties that may be encountered with any numerical root finding procedure are: moving into a region where the Jacobian  $\mathbf{J}_f$  is singular or nearly so; converging to a spurious solution, such as a minimum of  $r$  which is *not* a root of  $\mathbf{f}(\mathbf{e}) = \mathbf{0}$ ; and searching for roots whose order is greater than one, for example the root at  $x = 0$  of  $x^4 = 0$ . Since the current problem exists in a four dimensional space, the zero contours for this set of equations can not be plotted directly. Instead we will try to assess the nature of this root solving problem by plotting  $\sqrt{2}r$  (*i.e.*, the 2-norm of  $\mathbf{f}(\mathbf{e})$ ) versus two state variables at a time. This is reasonable since a root of  $\mathbf{f}(\mathbf{e}) = \mathbf{0}$  must be a minimum of  $r$ . These plots are shown in Figure 19 where Figure 19(a) shows the 2-norm of  $\mathbf{f}(\cdot)$  plotted with respect to  $x$  and  $y$  with  $z$  and  $\mathcal{S}$  held constant at 6.5 feet and  $2.5 \times 10^6$  counts per second respectively, and Figure 19(b) shows the 2-norm of  $\mathbf{f}(\cdot)$  plotted against  $z$  and  $\mathcal{S}$  with  $x$  and  $y$  fixed at 29 feet and 9 feet respectively. The detected count rates used to compute these two surfaces were obtained from the experiment described in Section 2 with the *Californium* source at the location labeled  $[3, (9, 0)]$  in Figure 3. This means that the actual source location on which these calculations were based was  $(29, 9, 6.5)$ . The source strength was not known exactly but was on the order of  $10^6$ . Note that the surfaces in Figure 19 were computed assuming a background level of 1200 counts per second, but it appears that the qualitative geometrical features are *unchanged* by altering this value. It is clear from Figure 19(a) that there is a well defined region where  $\|\mathbf{f}(x, y)\|_2$  is small. One difficulty is that all of the points on this horseshoe surrounding detector #3 have similar values of  $\|\mathbf{f}(x, y)\|_2$ , which may make finding the true minimum along this curve rather difficult numerically. Another potential problem revealed by this picture is that most of this surface is quite flat, meaning that it may be difficult for an algorithm to determine which direction



**Figure 19:** (a) The 2-norm of  $\mathbf{f}(\cdot)$  plotted with respect to  $x$  and  $y$  with  $z$  and  $\mathcal{S}$  held constant at 6.5 feet and  $2.5 \times 10^6$  counts per second respectively.  
(b) The 2-norm of  $\mathbf{f}(\cdot)$  plotted with respect to  $z$  and  $\mathcal{S}$  with  $x$  and  $y$  held constant at 29 feet and 9 feet respectively.

to move in order to find a minimum. Figure 19(b) also shows a well defined region where  $\|\mathbf{f}(z, \mathcal{S})\|_2$  is small, but it also presents the difficulty of finding the actual minimum along this parabolic curve of similar  $\|\mathbf{f}(z, \mathcal{S})\|_2$  values. Figure 19(b) implies that there are a series of source strength, source height pairs with very similar values of  $\|\mathbf{f}(z, \mathcal{S})\|_2$ , meaning it may be difficult to decide whether there is a stronger source high above or below the detector, or a weaker source level with the detector.

Three root solving methods have been implemented and tested on the source location functions given by Equations (11), (12), and (13). All three methods are local search techniques in the sense that information from a small neighborhood of the current position is used to decide which direction to move in order to get closer to the root. One of the algorithms used is the Newton-Raphson algorithm described previously. The Newton-Raphson method can diverge very rapidly, usually as a result of being near a horizontal point in the curve  $\mathbf{f}(\mathbf{e})$  (e.g., a minimum, maximum, or saddle point). This behavior is due to the fact that the derivative information used by the algorithm is all gathered at a single point  $\mathbf{e}_k$ . One way to correct this problem is to replace the derivative by the slope of the line segment connecting two different points,  $\mathbf{e}_k$  and  $\mathbf{e}_{k-1}$ , on the curve  $\mathbf{f}(\mathbf{e})$ . This modification is called the *secant method* in one dimension. One of the multi-dimensional generalizations is called *Broyden's method*, which is another of the algorithms used. Both the Newton-Raphson and Broyden's methods can stall if the curve  $\mathbf{f}(\mathbf{e})$  contains a region which is fairly flat (i.e., the Jacobian  $\mathbf{J}_f$  is close to being singular). If this occurs the search direction  $\delta \mathbf{e}$  may fail to exist entirely or may point away from a solution. One way to solve this problem is to consider a combination of two search directions, for instance one defined by  $-\mathbf{J}_f^{-1} \mathbf{f}$  (i.e., the Newton-Raphson search direction) and the other defined by  $\mathbf{J}_f^\dagger \mathbf{f}$  (i.e., the gradient  $\nabla r$  of  $\frac{1}{2} \mathbf{f}^\dagger \mathbf{f}$ ). The search direction used is the linear combination of these two directions which reduces  $r$  the most. The length of this step is restricted to be no longer than the length of the vector sum of the Newton and gradient vectors. A number of techniques of this sort exist, which collectively are known as *restricted step* or *trust region* methods. The final algorithm used is a trust region method due to *Powell*. Both the Newton-Raphson and Broyden's method implementations are modifications of the algorithms given in Press *et al.* (1992, Chapter 9). The implementation of Powell's method is a modification of the code in *minpack*, which is publicly available through the *NetLib* repository.

One detail of the root solving implementation that should be mentioned is that the equations which

are being solved can be written in three different ways,

$$\mathcal{S} - \frac{1}{\Omega_i(u_i, v_i, w_i) \epsilon_i \mathcal{F}_i} \left( \frac{\mathcal{D}_i(t)}{1 - \tau_i \mathcal{D}_i(t)} - \mathcal{B}_i(t) \right) = 0, \quad i = 1, 2, \dots, m, \quad (17a)$$

$$\mathcal{S} \Omega_i(u_i, v_i, w_i) - \frac{1}{\epsilon_i \mathcal{F}_i} \left( \frac{\mathcal{D}_i(t)}{1 - \tau_i \mathcal{D}_i(t)} - \mathcal{B}_i(t) \right) = 0, \quad i = 1, 2, \dots, m, \quad (17b)$$

$$\Omega_i(u_i, v_i, w_i) - \frac{1}{\mathcal{S} \epsilon_i \mathcal{F}_i} \left( \frac{\mathcal{D}_i(t)}{1 - \tau_i \mathcal{D}_i(t)} - \mathcal{B}_i(t) \right) = 0, \quad i = 1, 2, \dots, m. \quad (17c)$$

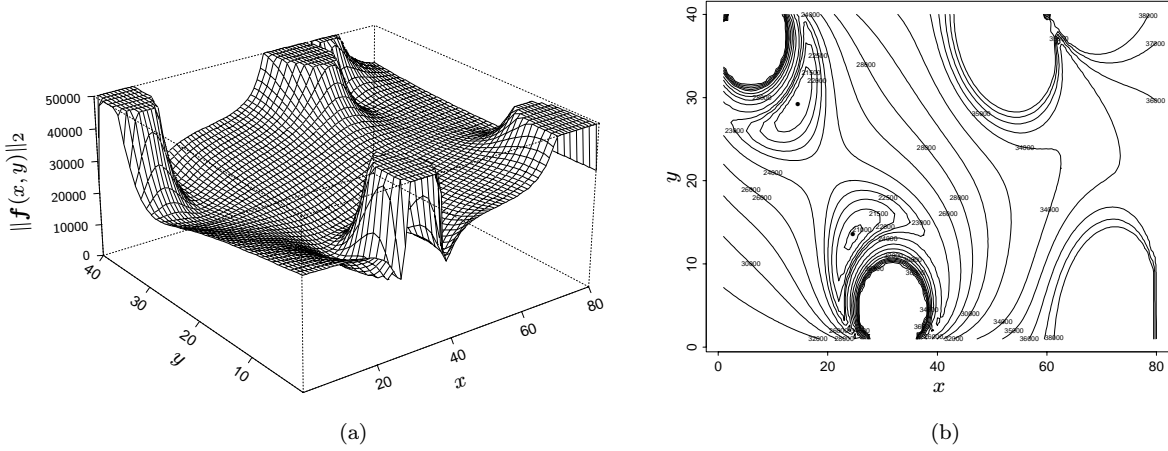
Each of the formulations in Equation (17) was tried with all three algorithms with the following results. None of the three methods converged to a solution located in the room when Equation (17c) was used. All three methods converged to reasonable solutions when Equations (17a) and (17b) were used. However, Equation (17b) yields a form of  $\frac{1}{2} \mathbf{f}^\dagger \mathbf{f}$  which approaches zero at the root, while Equation (17a) does not. For this reason Equation (17b) was used as the root function. The performance of these algorithms on the source location problem is shown in Table 4. In each of these runs the convergence

Algorithm Name	Number of Iterations	Final Residual $\ \mathbf{f}(\cdot)\ _2$	Final Solution
Newton-Raphson	70	44263	(39.7, 19.3, 8.7, $0.97 \times 10^6$ )
Newton-Raphson w/ Gradient Steps	112	3915	(28.8, 7.2, 8.6, $1.0 \times 10^6$ )
Broyden	10	44263	(39.7, 19.3, 8.7, $0.97 \times 10^6$ )
Broyden w/ Gradient Steps	83	3604	(27.1, 6.9, 8.6, $1.0 \times 10^6$ )
Powell	8	3518	(29.6, 7.0, 8.6, $1.0 \times 10^6$ )

**Table 4:** The performance of various instantiations of the three algorithms used to solve the source location problem.

tolerances were the same, and the algorithms were started at the same point, specifically the center of the room (40.0, 20.0, 4.0,  $1.0 \times 10^6$ ). Recall that the actual location of the source during this test was (29.0, 9.0, 6.5,  $\sim 10^6$ ). The background count level  $\mathcal{B}_i(t)$  was assumed to be 1200 counts per second. Changing the background level causes these algorithms to still converge to reasonable  $x$  and  $y$  values, but to unlikely the  $z$  and  $\mathcal{S}$  quantities. The algorithms denoted by the phrase w/ Gradient Steps take the first five steps in the direction of the gradient of  $r$  (e.g.,  $\mathbf{J}_f^\dagger \mathbf{f}$ ), rather than in the Newton direction (e.g.,  $-\mathbf{J}_f^{-1} \mathbf{f}$ ). Thereafter they use the Newton direction. Note that the algorithms which initially take gradient steps produce much better solutions than the corresponding algorithms without gradient steps. Also note that Powell's method is by far the best technique, since it produces the best solution, in the sense of the final residual value, in the fewest iterations. All of the algorithms have a slightly disturbing tendency to converge to the same height and source strength values. This reinforces the hypothesis that the U-shaped valley seen in Figure 19(b) consists of a series of height-source strength pairs which all have very similar residual values. It remains unclear why a height of about 8.6 feet acts as an attractor for all three algorithms.

As a final experiment we will try to assess the difficulty of locating two sources using Equation (11). In order to do this, detector readings were artificially constructed from the data for one source. This was done by adding the detected count rates for the one source case to themselves with the detector rotated one step clockwise and dividing the sum by two. In other words, the new count rate at detector #3 was created by adding the old rate at #3 to the old rate at #6 and dividing by two. Likewise the new rate at #5 was computed by averaging the old rates at #5 and #3, etc. The effect of this is to

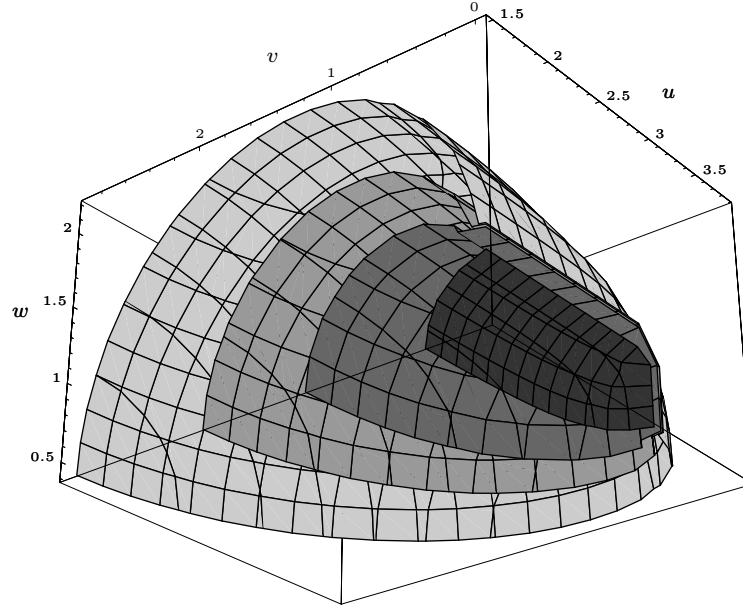


**Figure 20:** (a) The 2-norm of  $\mathbf{f}(\cdot)$  plotted with respect to  $x$  and  $y$  with  $z$  and  $\mathcal{S}$  held constant at 6.5 feet and  $2.5 \times 10^6$  counts per second respectively. Note that there two sources, one in front of sensor #3, the other in front of sensor #5. (b) The surfaces of constant value of the 2-norm of  $\mathbf{f}(\cdot)$ . Note the two *separate* minima in the residual surface, each marked with a black dot  $\bullet$ .

simulate the count rates for two similar strength sources, one 9 feet in front of sensor #3, the other 9 feet in front of sensor #5. The resulting residual surface is shown in Figure 20(a). The surfaces of constant value (*i.e.*, level surfaces) of this plot are shown in Figure 20(b). The black dots show the two separate minima in this surface that correspond to the two sources. Note that the residual values  $\|\mathbf{f}(\cdot)\|_2$  at these two minima are within 5% of each other. Figure 20 indicates that finding multiple sources may be a tractable problem if the sources are each relatively close to a different sensor and if a root finding algorithm which finds *all* roots can be developed. One practical difficulty associated with tracking multiple sources using root finding is that there must be exactly four detectors for each source in order to have the same number of equations as unknowns. This scaling difficulty motivates the use of a more general method which can be more easily adapted to a multi-source problem.

### 5.3 Locating Sources using Least Squares Estimation

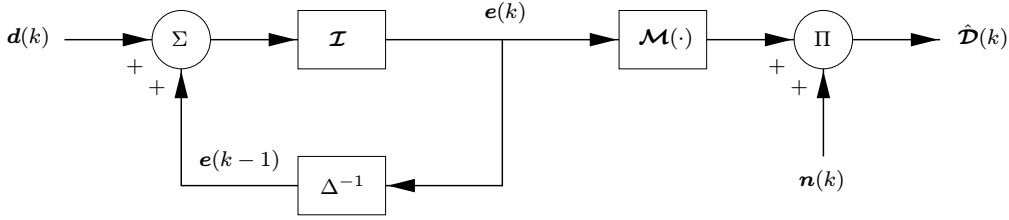
In this section we discuss an estimation algorithm used to track the position of the radioactive source over time. Note that in principle this least squares estimation method can be readily extended to tracking multiple sources, unlike the root finding method. We begin by discussing how to use the model in Equation (11) to locate a radioactive source in this framework. Assume for the moment that the detected count rate  $\mathcal{D}_i$  obeys the deterministic model in Equation (11). For a single source, the count rate  $\mathcal{D}_i$  at the  $i$ th detector defines a single level curve of  $\mathcal{M}_i(u, v, w)$  given by  $\mathcal{M}_i(u, v, w) = \mathcal{D}_i$ . There is one such curve for each detector and their intersection points define the possible source locations. The level curves for a single detector are surfaces in 3-dimensions which are illustrated in Figure 21. The four surfaces in the figure represent solid angles  $\Omega_i(u, v, w)$  of  $\pi$ ,  $\frac{\pi}{3}$ ,  $\frac{\pi}{6}$ , and  $\frac{\pi}{12}$ , with  $\pi$  being the innermost (*i.e.*, darkest color) surface, and  $\frac{\pi}{12}$  being the outermost (*i.e.*, lightest color) surface. The cut-away shown is taken through the center of the detector, and due to symmetry the whole surfaces can be constructed by appropriate reflections of this figure around that center point. Therefore the complete surfaces are approximately ellipsoids, except for a neighborhood near the edges of the detector. Since the count rate is assumed to be consistent with the model, these four surfaces intersect at only one point,



**Figure 21:** The level surfaces for a single detector corresponding to the model in Equation (11). The four surfaces shown correspond to solid angles  $\Omega_i(u, v, w)$  of  $\pi$  (darkest color),  $\frac{\pi}{3}$ ,  $\frac{\pi}{6}$ , and  $\frac{\pi}{12}$  (lightest color).

which corresponds to the source location. Now assume that the count rate  $\mathcal{D}_i$  is randomly distributed with a mean  $\mathcal{M}$  given by the model in Equation (11) and a variance  $\mathcal{V}$  proportional to the mean. In this case the level surfaces become shells whose thickness is proportional to  $\sqrt{\mathcal{V}}$ , and their intersection becomes some volume element in 3-dimensions wherein the source is located. So in this stochastic setting, the problem is to find the most likely location for a given set of detected count rates.

This problem was posed as a state estimation problem whose formulation is explained, for instance, in Sage and Melsa (1971, Chapter 9), Jazwinski (1970, Chapter 6), and Bryson and Ho (1975, Chapter 12). In this context the measurements or outputs are the detected count rate  $\mathcal{D}_i(k)$  at each detector  $i$  at each time step  $k$ . The states are chosen according to the problem that must be solved. In this case, the state must at least contain the source location  $(x(k), y(k), z(k))$  at every time step  $k$ . In addition to the source location, another unknown in Equation (11) is the background count rate  $\mathcal{B}_i$  at each detector. We assume that the background is the same at each detector, making  $\mathcal{B}(k)$  independent of  $i$ , and include this single quantity in the state. Denote any change in the state estimates at time  $k$  by the vector  $\mathbf{d}(k) = [\delta x(k) \ \delta y(k) \ \delta z(k) \ \delta \mathcal{B}(k)]^\dagger$ , denote the state estimates at a particular time step  $k$  by the vector  $\mathbf{e}(k) = [x(k) \ y(k) \ z(k) \ \mathcal{B}(k)]^\dagger$  and call the number of elements in each of these vectors  $p = 4$ . We chose to approximate the detected count rates with the vector  $\hat{\mathcal{D}}(k)$  drawn from a Poisson process with mean vector  $\mathcal{M}(\mathbf{e}(k))$ , where  $\mathbf{e}(k)$  is the current state estimate. The structure for our approximate system is shown in Figure 22. In this figure,  $\mathcal{M}(\cdot)$  is the vector form for the right hand side of Equation (11),  $\mathcal{I}$  is the identity mapping,  $\Delta^{-1}$  is a unit time delay, and  $\mathbf{n}(k)$  is a multiplicative noise process. Multiplicative noise is chosen because the *link function* defined in McCullagh and Nelder (1989, Chapter 6) for a Poisson distribution is  $\exp(\cdot)$ . Under this mapping the expected value for the approximate count rate  $\hat{\mathcal{D}}_i(k)$  is  $\mathcal{E}\{\hat{\mathcal{D}}_i(k)\} = \exp(\mathcal{E}\{\tilde{\mathcal{M}}_i\}) \exp(\mathcal{E}\{n_i\})$ . Define the deterministic part of the model by  $\exp(\mathcal{E}\{\tilde{\mathcal{M}}_i\}) = \mathcal{M}_i(\mathbf{e}(k))$ , and let the noise portion  $n_i(k)$  be a zero mean process. Therefore the noise mean is  $\exp(\mathcal{E}\{n_i\}) = 1$ , and the modeled mean count rate is  $\mathcal{E}\{\hat{\mathcal{D}}_i(k)\} = \mathcal{M}_i(\mathbf{e}(k))$ . Since we want  $\hat{\mathcal{D}}_i(k)$  to have the properties of a Poisson process, its variance must also be  $\mathcal{M}_i(\mathbf{e}(k))$ . This will be the case if the noise process is chosen such that  $\exp(\mathcal{E}\{n_i^2\}) = \mathcal{M}_i(\mathbf{e}(k))$ . This means that



**Figure 22:** A block diagram of the system approximation.

the characteristics of the stochastic portion of the model are *dependent* on the deterministic part of the model, in marked contrast to typical system models with Gaussian noise.

We find the maximum a posteriori (MAP) estimate for the state at each time step. As pointed out in Sage and Melsa (1971, Chapter 9), for a Gaussian noise process this is equivalent to finding a weighted least squares fit to the measurements. This turns out to be true even when the system is nonlinear. For Gaussian noise the proper weights are the noise covariance matrices. Based on the analysis in McCullagh and Nelder (1989, Chapter 2), we will show that for a Poisson noise process, the MAP estimate is still equivalent to a weighted least squares fit, but with different weights. We have chosen the MAP estimate over the conditional mean (*i.e.*, minimum variance) or conditional median (*i.e.*, minimum absolute error) estimates because we must incorporate state constraints into the estimation process. This turns out to be quite easy if one is using weighted least squares, even when the model is nonlinear. Note that we did not choose an extended Kalman filter as an estimation method in our problem both because we have state constraints, and because the noise process in our problem is Poisson rather than Gaussian. Furthermore, a study in Jazwinski (1970) concludes that for systems with relatively large measurement nonlinearities, such as ours, the extended Kalman filter converges rather slowly to the actual state.

We start our estimation process at time  $k = 0$  with an estimate of the source position and background level  $(x(0), y(0), z(0), \mathcal{B}(0))$ . We assume that this initial estimate consists of the correct values at this time. We also assume that the absolute source strength  $\mathcal{S}$  is known, and that the correction factor product  $\mathcal{F}_i$  is constant over time. Note that these assumptions may not always be reasonable, but in our problem they are often valid. At succeeding time steps we update our estimate of the source location and background level using the relations

$$\begin{aligned} x(k) &= x(k-1) + \delta x^*(k), \\ y(k) &= y(k-1) + \delta y^*(k), \\ z(k) &= z(k-1) + \delta z^*(k), \\ \mathcal{B}(k) &= \mathcal{B}(k-1) + \delta \mathcal{B}^*(k), \end{aligned} \tag{18}$$

for all time steps  $k = 1, 2, 3, \dots$ , where  $\delta x^*(k)$ ,  $\delta y^*(k)$ ,  $\delta z^*(k)$ , and  $\delta \mathcal{B}^*(k)$  are state changes which are optimal in some sense. Note that using our previous definitions, Equation (18) can be written in vector form as  $\mathbf{e}(k) = \mathbf{e}(k-1) + \mathbf{d}^*(k)$ .

At each time step  $k$  the optimal estimated state change  $\mathbf{d}^*(k)$  is computed by solving the optimization problem

$$\begin{aligned} \mathbf{d}^*(k) &= \arg \min_{\mathbf{d}(k)} \left( \sum_{i=1}^m \mathcal{W}_i^m \left( \mathcal{D}_i(k) - \mathcal{M}_i(\mathbf{e}(k-1) + \mathbf{d}(k)) \right)^2 + \sum_{j=1}^p \mathcal{W}_j^d d_j^2(k) \right), \\ \text{subject to} & \\ -\mathcal{N}_j &\leq d_j(k) \leq \mathcal{N}_j, \\ \mathcal{L}_j &\leq e_j(k-1) + d_j(k) \leq \mathcal{U}_j, \quad j = 1, \dots, p, \quad k = 1, 2, \dots \end{aligned} \tag{19}$$

In this equation,  $\mathcal{M}_i(\mathbf{e}(k-1) + \mathbf{d}(k))$  represents the mean value of our approximation of the detected

count rate  $\mathcal{E}\{\hat{\mathcal{D}}_i\}$ , given by the right hand side of Equation (11) evaluated at  $\mathbf{e}(k-1) + \mathbf{d}(k)$  for the  $i$ th detector. Note that evaluating the right hand side of Equation (11) requires solving Equations (12) and (13) first. The quantity  $\mathcal{W}_i^m$  is a weight which determines how closely the algorithm tries to match the model. Similarly,  $\mathcal{W}_j^d$  sets how much the algorithm is allowed to change the previous state estimate. The term  $\sum_{i=1}^m \mathcal{W}_i^m (\mathcal{D}_i(k) - \mathcal{M}_i(\mathbf{e}(k-1) + \mathbf{d}(k)))^2$  is a weighted  $L_2$  norm of the difference  $\|\mathcal{D}(k) - \mathcal{E}\{\hat{\mathcal{D}}(k)\}\|_2$  between the actual detected count rate  $\mathcal{D}(k)$  and the expected value of the predicted detected count rate  $\mathcal{E}\{\hat{\mathcal{D}}(k)\}$ . This term corresponds to a weighted least squares fit of the state estimate to the measurements. Likewise the term  $\sum_{j=1}^p \mathcal{W}_j^d d_j^2(k)$  is the weighted  $L_2$  norm of the difference  $\|\mathbf{e}(k) - \mathbf{e}(k-1)\|_2$  between the estimated position and background at the previous time and that at the current time. This is a regularization term that penalizes large changes in the state estimate more than small changes. So Equation (19) will try to minimize the change in position and background while simultaneously making a change that makes the resulting position and background approximately agree with the model. When selecting the weights, note that making  $\mathcal{W}_i^m$  large relative to  $\mathcal{W}_j^d$  reflects a high confidence in the sensor model, while setting  $\mathcal{W}_j^d$  high with respect to  $\mathcal{W}_i^m$  indicates confidence in the sensor measurements.

As stated previously, the MAP state estimate for a Poisson noise process is equivalent to a weighted least squares fit of the measurements with the appropriate weights. Maximizing the a posteriori probability  $\mathcal{P}\{\mathbf{e}(k) | \mathcal{D}(k)\}$  is equivalent by Bayes rule to maximizing the probability  $\mathcal{P}\{\mathcal{D}(k) | \mathbf{e}(k)\} \mathcal{P}\{\mathbf{e}(k)\}$ . Assuming that the measurements at a given time step  $\mathcal{D}(k)$  are statistically independent, then given the source location  $\mathbf{e}(k)$ , the conditional probability for the measurements  $\mathcal{D}(k)$  is simply the product of Poisson distributions

$$\mathcal{P}\{\mathcal{D}(k) | \mathbf{e}(k)\} = \prod_{i=1}^m \exp\left(\frac{1}{\gamma_i^2} \mathcal{D}_i(k) \ln(\mathcal{M}_i(\mathbf{e}(k))) - \frac{1}{\gamma_i^2} \mathcal{M}_i(\mathbf{e}(k)) - \ln(\mathcal{D}_i(k)!)\right), \quad (20)$$

in which  $\mathcal{M}_i(\mathbf{e}(k))$  is the mean count rate for the  $i$ th detector given the state  $\mathbf{e}(k)$ . Each component of this Poisson distribution has a variance equal to  $\gamma_i^2$  times the mean,  $\mathcal{V}(\mathcal{D}_i) = \gamma_i^2 \mathcal{M}_i(\mathbf{e}(k))$ . Assuming that all states  $\mathbf{e}(k)$  are equally likely, we chose a uniform a priori probability  $\mathcal{P}\{\mathbf{e}(k)\}$ . The uniform a priori probability only scales the conditional probability  $\mathcal{P}\{\mathcal{D}(k) | \mathbf{e}(k)\}$ , and so does not effect the locations of extrema. Furthermore, the extremal points are invariant under the monotonic transformation  $\ln(\cdot)$ . Hence we will maximize the natural logarithm of the quantity in Equation (20), which we call the *likelihood function*  $\mathcal{L}(\mathcal{D}; \mathcal{M})$ . Maximizing  $\mathcal{L}(\mathcal{D}; \mathcal{M})$  is equivalent to minimizing the difference between the largest possible value for  $\mathcal{L}(\cdot)$ , which is when  $\mathcal{M} = \mathcal{D}$ , and the value of  $\mathcal{L}(\cdot)$  for the estimate  $\mathcal{M}$ . All of this discussion implies that the MAP estimate for  $\mathbf{e}(k)$  under a Poisson distribution is obtained by minimizing the quantity

$$2 \sum_{i=1}^m \frac{1}{\gamma_i^2} \left( \mathcal{D}_i(k) \left( \ln \mathcal{D}_i(k) - \ln(\mathcal{M}_i(\mathbf{e}(k))) \right) - \left( \mathcal{D}_i(k) - \mathcal{M}_i(\mathbf{e}(k)) \right) \right). \quad (21)$$

This result is analogous to the one obtained in McCullagh and Nelder (1989, Chapter 6) for a univariate Poisson distribution. An equivalence can be established between the *first* term in Equation (19) and Equation (21) by taking the first order Taylor series expansion of  $\ln \mathcal{D}_i(k)$  about the point  $\mathcal{M}_i(\mathbf{e}(k))$ , and substituting this expansion into Equation (21). It is straightforward to show that the result of this substitution is identical to the first term in Equation (19) if the model weights are chosen as  $\mathcal{W}_i^m = \frac{1}{\gamma_i^2 \mathcal{M}_i(\mathbf{e}(k))}$ . Note that this choice of weights results in the diagonal weight matrix  $\mathcal{W}^m = \text{diag}_{i=1, \dots, m} ((\gamma_i^2 \mathcal{M}_i(\mathbf{e}(k)))^{-1})$ . The quantity  $\gamma_i^2$  is the constant of proportionality between the variance and the mean, called the *dispersion*. It is stated in McCullagh and Nelder (1989, Chapter 6) that in practice  $\gamma_i^2$  is often greater than 1. For example, this occurs when samples of the Poisson process are made at random time intervals, rather than fixed intervals. This occurs in our problem because the length of time between count rate samples  $\mathcal{D}_i(k)$  for each detector varies over time. Furthermore, the order in which the different detectors are sampled varies with time. We will discuss our choice for the dispersion parameters  $\gamma_i^2$  in Section 5.4.

Note that there are upper and lower bounds on the state change  $\mathbf{d}(k)$  and on the state estimate  $\mathbf{e}(k)$ . The constraints on the estimate  $\mathbf{e}(k)$  are imposed so that the estimated position can not be outside the room, and the estimated background can not exceed historical bounds for the background levels in the room. The constraints on the change in estimate  $\mathbf{d}(k)$  are imposed in order to keep noise from causing the position to jump around excessively from time step to time step. Conceptually these constraints are reasonable because the sources are moved by people, and there is a limit to how far a person can move in a given time interval.

The optimization problem posed in Equation (19) is a nonlinear programming problem with linear inequality constraints, which can be solved by a number of methods, many of which are discussed in Bertsekas (1995, Chapter 4) and Fletcher (1987, Chapter 12). The optimization algorithm that we chose was CFSQP, which is documented in Lawrence, Zhou, and Tits (1994). This algorithm uses a sequential quadratic programming (SQP) approach, modified so that each iteration is feasible with respect to the constraints. This algorithm is designed to minimize the maximum of a set of objective functions (*i.e.*, *minimax* problems), but since our set of objective functions contains only one member it also solves the problem posed in Equation (19). We chose this algorithm because all the intermediate iterates of the algorithm are feasible, so if we have to stop the optimization before achieving convergence, the resulting suboptimal solution will still satisfy the constraints. This is an important consideration in our application because we need to track the sources in real-time relative to the average detector sampling time, which in our case is 1 second. Hence if our optimization runs for more than 1 second at time step  $k$  we simply stop it, use the resulting suboptimal solution for the estimate  $\mathbf{e}(k)$ , and then begin estimating the next position  $\mathbf{e}(k+1)$  with the new sensor data.

The method outlined so far computes the current state estimate  $\mathbf{e}(k)$  based only on the current measurement  $\mathcal{D}(k)$ . In principle a better estimate may be obtained if the last  $l$  measurements ( $\mathcal{D}(k), \mathcal{D}(k-1), \dots, \mathcal{D}(k-(l-1))$ ) are used. This is the idea behind the moving horizon state estimation procedures described in Muske and Rawlings (1995). The optimal state change  $\tilde{\mathbf{d}}^*(k)$  over the entire window is found by solving the optimization problem

$$\tilde{\mathbf{d}}^*(k) = \arg \min_{\tilde{\mathbf{d}}(k)} \left( \sum_{n=0}^{l-1} \left( \sum_{i=1}^m \mathcal{W}_i^m \left( \mathcal{D}_i(k-n) - \mathcal{M}_i(\mathbf{e}(k-n-1) + \mathbf{d}(k-n)) \right) \right)^2 + \sum_{j=1}^p \mathcal{W}_j^d d_j^2(k-n) \right), \quad (22)$$

subject to

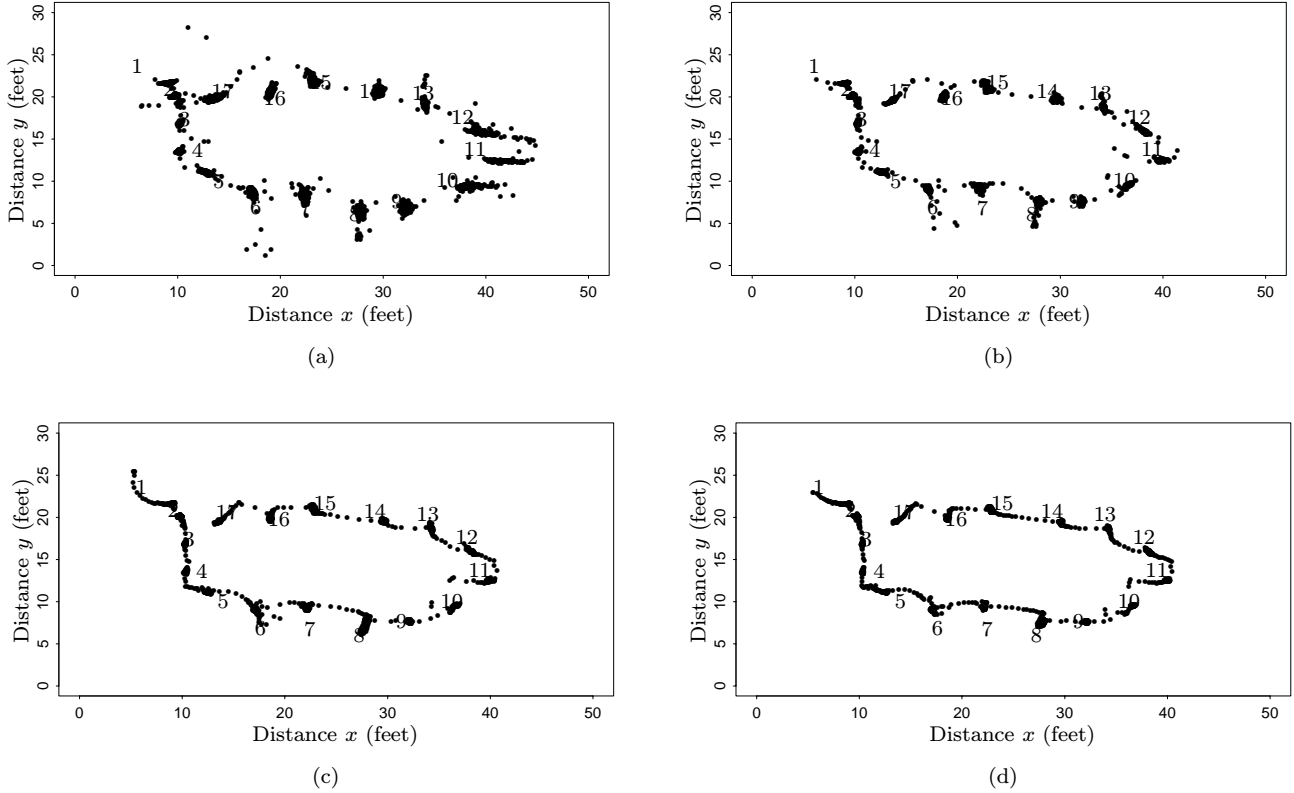
$$\begin{aligned} -\mathcal{N}_j &\leq d_j(k-n) \leq \mathcal{N}_j, \\ \mathcal{L}_j &\leq e_j(k-n-1) + d_j(k-n) \leq \mathcal{U}_j, \quad j = 1, \dots, p, \quad n = 0, \dots, l-1, \quad k = 1, 2, \dots \end{aligned}$$

This equation reduces to Equation (19) when the window length  $l$  is equal to 1. In Equation (22) the state change for the entire window at time step  $k$  is  $\tilde{\mathbf{d}}(k) = [\mathbf{d}(k) \mathbf{d}(k-1) \dots \mathbf{d}(k-(l-1))]^\dagger$ , and the state estimate is  $\tilde{\mathbf{e}}(k) = [\mathbf{e}(k) \mathbf{e}(k-1) \dots \mathbf{e}(k-(l-1))]^\dagger$ . This algorithm is initialized by performing a batch estimation until sample time  $l$ . Muske and Rawlings (1995) present examples from several application areas in which this method outperforms one-step (*i.e.*,  $l = 1$ ) recursive estimation.

## 5.4 Tracking Results

In this section we present tracking results based on experiments with a source in the facility. Note that the experiment used to test the source tracking is *different* from the experiment used for sensor characterization. The experimental procedure consisted of placing a  $\text{Cs}^{137}$  source on a cart and moving it along a predetermined path in the room. Along this path there were 17 points at which we paused with the source for approximately 3 minutes each. The purpose of the 3 minute pauses was to collect enough

data at each position to compute a mean count rate for each sensor at each position. The resulting means are used to compare the performance of our algorithm on the actual data to its performance on idealized Poisson data. The estimated position of the source as seen from the ceiling looking down is shown in Figure 23. Figure 23(a) shows the estimated source positions for dispersion  $\gamma_i^2 = 1$  for all

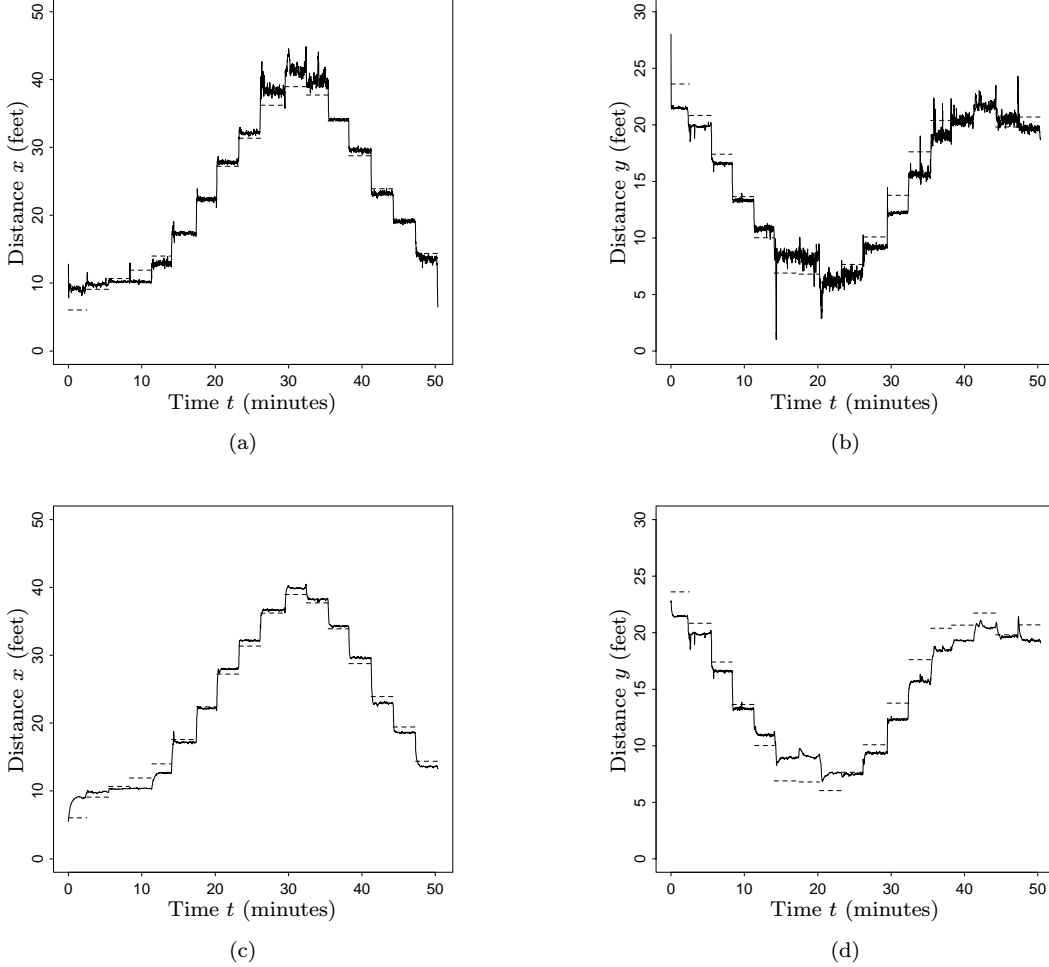


**Figure 23:** The estimated position of the source in the  $x$ - $y$  plane with dispersion (a)  $\gamma_i^2 = 1$  (b)  $\gamma_i^2 = 10$  (c)  $\gamma_i^2 = 50$  (d)  $\gamma_i^2 = 100$  for window length  $l = 1$ . These plots show a view of the room from the ceiling looking down. The detectors are located in each of the four corners of these pictures.

$i$ , Figure 23(b) shows the locations with  $\gamma_i^2 = 10$ , Figure 23(c) illustrates positions for  $\gamma_i^2 = 50$ , and Figure 23(d) gives estimates for  $\gamma_i^2 = 100$ . In all cases the state weights are  $\mathcal{W}_j^d = 1$  and the window length is  $l = 1$ . The 17 numbers in these plots show the actual positions of the points at which we paused with the source. Beginning at point 1, we followed the numbered points in ascending order, ending at point 17. The points  $\bullet$  in the figure represent the pairs  $(x(k), y(k))$  of the state estimate for every time step  $k$ . These estimates are computed at 1 second intervals, which is also the average sampling time for the detectors. Our algorithm runs fast enough for us to compute each state estimate in less than 1 second, hence we are able to track the source in real-time relative to the average sampling rate of the detector. Although the detectors do not appear in these pictures, they are located in each corner at a  $45^\circ$  angle with respect to the walls. Note that as the dispersion  $\gamma_i^2$  is increased from Figure 23(a) to Figure 23(d), the position estimates go from clusters around the 17 numbered points, to a smooth track running through these 17 points.

Figure 23 shows that most of the estimates of the source location are close to the path described by the 17 points where we paused with the source. However, it does not show whether the estimates follow this path in the correct temporal order. Figure 24 shows the estimated positions in both the  $x$

and  $y$  directions versus time for the  $x$ - $y$  plots shown in Figures 23(a) and 23(d). In these two plots the



**Figure 24:** The estimated positions of the source for dispersion  $\gamma_i^2 = 1$  in the (a)  $x$ -direction (b)  $y$ -direction versus time. The calculated source locations for dispersion  $\gamma_i^2 = 100$  in the (c)  $x$ -direction (d)  $y$ -direction versus time. The dotted lines show the actual source positions over time.

the solid lines represent the estimated source position, and the dotted lines represent the actual source position. Figures 24(a) and 24(c) plot the positions in the  $x$ -direction versus time for dispersions  $\gamma_i^2 = 1$  and  $\gamma_i^2 = 100$  respectively. Likewise Figures 24(b) and 24(d) show the  $y$ -position with respect to time for the same two dispersions. It is clear from these pictures that the numbered points are visited in the correct order.

We estimated the error of this tracking algorithm using the following two statistics. For both statistics we begin by taking the absolute value of the difference between the median estimated position and the actual position for each of the 17 positions in both the  $x$  and  $y$  directions. This corresponds to taking the median value of each plateau in Figure 24, subtracting each median from each dotted line, and taking the absolute value of each result. This results in two sets of 17 numbers, one for each direction  $x$  and  $y$ . For the first statistic, combine the elements within each set by taking the maximum value over the set. Call the resulting maximum absolute error in the  $x$ -direction  $\mathcal{E}_x^{max}$ , and in the  $y$ -direction

call it  $\mathcal{E}_y^{max}$ . For the second statistic, take the median value to combine the differences within each set. Denote the median absolute error along  $x$  by  $\mathcal{E}_x^{med}$  and along  $y$  by  $\mathcal{E}_y^{med}$ . These two statistics are then used to compute two percent location error statistics for the algorithm. To find the maximum percent location error  $\mathcal{R}_e^{max}$ , construct a rectangle of width  $2\mathcal{E}_x^{max}$  and height  $2\mathcal{E}_y^{max}$ . Compute the percent location error  $\mathcal{R}_e^{max}$  by dividing the area of this rectangle by the area of the room. Similarly the median percent location error  $\mathcal{R}_e^{med}$  is computed by dividing the area of the  $2\mathcal{E}_x^{med} \times 2\mathcal{E}_y^{med}$  rectangle by the room area. For the dispersion values  $\gamma_i^2 = 1$  and  $\gamma_i^2 = 100$  shown in Figure 24, the values of these error statistics are given in Table 5. Based on these two error statistics it appears that the lower

$\gamma_i^2$	$\mathcal{E}_x^{max}$ (feet)	$\mathcal{E}_y^{max}$ (feet)	$\mathcal{R}_e^{max}$ (%)	$\mathcal{E}_x^{med}$ (feet)	$\mathcal{E}_y^{med}$ (feet)	$\mathcal{R}_e^{med}$ (%)
1	3.137	2.141	2.117	0.740	0.909	0.212
100	2.916	2.284	2.099	0.769	1.358	0.329

**Table 5:** Two error measures for the location estimates shown in Figure 24.

dispersion value  $\gamma_i^2 = 1$  leads to a more accurate estimate than the higher dispersion value  $\gamma_i^2 = 100$ . Also it appears that increasing the dispersion increases the time lag between the estimated and actual locations.

The statistics in Table 5 become more meaningful when compared with the those from an idealized situation. As a baseline against which to evaluate our performance, we chose a scenario in which the gamma emissions from the source were assumed to be a Poisson process with dispersion  $\gamma_i^2 = 1$ , and the detectors were assumed to output the exact gamma count rate with no additional noise. To test the algorithm on this ideal scenario we generated synthetic data drawn from a Poisson process as the output of each detector. This synthetic experiment was identical to our actual experiment in that the synthetic source paused for the same length of time at the same 17 positions. The synthetic data was generated by first taking the average  $\mu_l^i$  of the actual count rate for each detector  $i = 1, \dots, 4$  at each position  $l = 1, \dots, 17$ . We then computed the number of measurements  $s_l$  made at each position  $l = 1, \dots, 17$ . Using these two numbers we drew  $s_l$  samples from a Poisson distribution with mean  $\mu_l^i$  for all detectors  $i = 1, \dots, 4$  and all positions  $l = 1, \dots, 17$ . Twenty-five synthetic data sets were created in this fashion, and our algorithm was run on each of them with a window length of  $l = 1$ . We then computed the median absolute error in both the  $x$  and  $y$  directions for the estimates generated with each of these twenty-five synthetic data sets. For comparison to the results with actual data, we calculated the mean and standard deviation of this statistic over the twenty-five synthetic runs. The results are presented in Table 6. Notice that both  $\mathcal{E}_x^{med}$  and  $\mathcal{E}_y^{med}$  for the actual data fall within one standard deviation of the

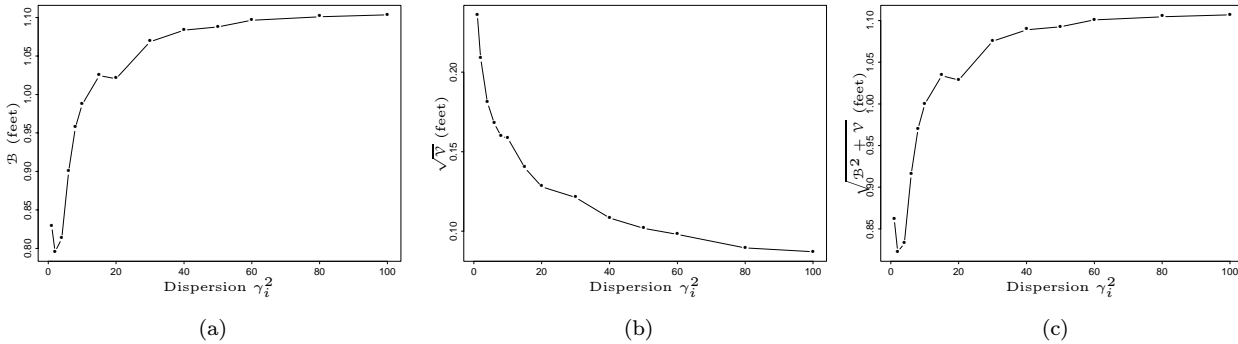
	Mean $\mathcal{E}_x^{med}$ (feet)	Deviation $\mathcal{E}_x^{med}$ (feet)	Mean $\mathcal{E}_y^{med}$ (feet)	Deviation $\mathcal{E}_y^{med}$ (feet)
Actual	0.740	—	0.909	—
Synthetic	0.758	0.0219	0.922	0.0156

**Table 6:** A comparison of the median absolute error between the actual data and synthetic data drawn from a Poisson distribution.

means for the synthetic data. We conducted a Wilcoxon test, described for instance in Conover (1980, Chapter 5) and Hollander and Wolfe (1973, Chapter 2), of the hypothesis that the value of  $\mathcal{E}_x^{med}$  for the actual data was drawn from a distribution having the same mean as the one for the synthetic data. We conducted a similar hypothesis test for the statistic  $\mathcal{E}_y^{med}$ . These hypothesis tests gave  $p$ -values of 0.0006 and 0.0003 for  $\mathcal{E}_x^{med}$  and  $\mathcal{E}_y^{med}$  respectively. The  $p$ -value specifies the *maximum* significance at which the null hypothesis should be accepted. So if the desired level of significance is less than the  $p$ -value, accept the null hypothesis, otherwise reject it. The level of significance is the probability of accepting

the alternative hypothesis when the null hypothesis is true, so it is conceptually identical to a maximum false alarm rate. We chose 0.01 (*i.e.*, a 1% false alarm probability) as a reasonable significance, therefore with 99% confidence both  $\mathcal{E}_x^{med}$  and  $\mathcal{E}_y^{med}$  for the actual data came from distributions with means that were *different* from the synthetic data distributions. Note that the closer the  $p$ -value is to one, the more likely it is that the means of the actual distributions are equal to the means of the synthetic distributions. Several conclusions emerge from this analysis. First, the performance of our algorithm on real data is qualitatively very similar to its performance on idealized synthetic data, which indicates that our model captures the major features of the real data. Second, since the errors for the actual data have a different mean than those for the synthetic data, our assumption that the dispersion  $\gamma_i^2$  is one is probably incorrect. Third, since the median absolute errors for both the actual and synthetic data are not zero, there is probably a minor systematic error in our sensor model. We conjecture that the primary source of unmodeled error comes from neglecting the change in gamma absorption with position due to the finite detector thickness, but we have not attempted to model this phenomena.

We now address the issue of choosing an appropriate value for the dispersion. It is clear from Figures 23 and 24 that the deviation between the estimated and actual position, called the *bias*, changes as the dispersion  $\gamma_i^2$  changes. The estimator bias is given mathematically by the expression  $\mathcal{B} = |\mathcal{E}\{e\} - \mathcal{E}\{p\}|$ , where  $\mathcal{E}\{p\}$  is the true source position and background count rate. Note that this definition means that our error statistics from the previous paragraph are the maximum model bias and the median model bias as a percentage of the room size. It can also be seen in Figures 23 and 24 that the deviation between the estimated position and the average estimated position, called the *variance*, also changes with dispersion. The estimator variance is given by  $\mathcal{V} = \mathcal{E}\{(e - \mathcal{E}\{e\})^2\}$ . One way to choose the dispersion parameter  $\gamma_i^2$  is by selecting the value which minimizes the sum of the bias squared and the variance. Since the sum of squared bias and variance represents the total error made by the model, this choice of dispersion will minimize the total error. Plots of bias, standard deviation, and total error for the  $x$ - $y$  plane for a range of dispersion values are shown in Figure 25. Figure 25(a) plots the estimator



**Figure 25:** The change in (a) estimator bias (b) estimator standard deviation (c) total estimator error as a function of the dispersion parameter  $\gamma_i^2$ .

bias versus dispersion  $\gamma_i^2$ , Figure 25(b) shows the change in standard deviation with dispersion, and Figure 25(c) illustrates the variation in total estimator error as dispersion changes. Comparing the relative sizes of the errors in Figures 25(a) and 25(b), it is clear that bias is the dominant error source in this problem. Also note that bias generally increases as dispersion increases, while standard deviation decreases with increasing dispersion. This seems reasonable if the dispersion is considered a smoothness parameter, since increasing the smoothness tends to decrease standard deviation and increase bias. From Figure 25(c) it appears that the minimum value for the total error occurs at the dispersion  $\gamma_i^2 = 2$ . Analyzing the variance to mean ratio of the data gives an average dispersion value of 1.60, which is in excellent agreement with Figure 25(c).

We found that with the system model shown in Figure 22, the state estimates for all window lengths

$l > 1$  were almost identical to the estimates for window length  $l = 1$ . We believe that this observation can be explained by the fact that the model in Figure 22 has no dynamics, hence any change in the estimate  $\mathbf{e}(k-1)$  is completely determined by the external driving force  $\mathbf{d}(k)$ . While it may appear that the regularization term  $\|\mathbf{e}(k) - \mathbf{e}(k-1)\|_2$  provides some coupling between the states at different times, recall that the current state estimate  $\mathbf{e}(k)$  is selected so as to minimize this term. Therefore the observation that there is only very weak coupling between states at different times seems reasonable. If the model contained terms approximating the source velocities, then perhaps a moving horizon state estimate with window length  $l > 1$  would be different from a state estimate with window length  $l = 1$ . However, velocity estimation has questionable value in this problem because a person can change velocities much more quickly than our average detector sampling rate.

## 6 Conclusion

We have described real-time algorithms for both detecting the presence and tracking the position of radioactive sources in a facility in the presence of measurement noise. We formulated the detection problem as a nonparametric hypothesis testing problem. This problem is solved by comparing a statistic computed over some window(s) of the data to a threshold value. If this threshold is exceeded then a source is present. We formulated the tracking problem as a state estimation problem and solved it recursively using a constrained nonlinear optimization method. The optimization simultaneously minimizes the change in source position and disagreement between measurements and a sensor model. The sensor model is a fairly complex function relating position to detected count rate.

The overall purpose of this work is to enhance both security and safety by automating part of the assessment process, allowing remote assessment, and introducing new sensor modalities into the assessment process. We presented detection and tracking results based on experiments done with one source in a single room. Our results indicate that a source can be detected and tracked quite well with these algorithms in spite of fairly poor signal to noise ratios, and rather high measurement noise levels. In short, we have demonstrated the capability to detect and track a single source in real-time with high accuracy in spite of a complex sensor model, an unknown background signal, and high measurement noise.

## References

- Basseville, M., & Nikiforov, I. (1993). *Detection of Abrupt Changes: Theory and Application*. Information and Systems Science Series. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Bertsekas, D. (1995). *Nonlinear Programming*. Optimization and Neural Computation Series. Athena Scientific, Belmont, MA.
- Box, G., Jenkins, G., & Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control* (3rd edition). Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Bryson, A., & Ho, Y.-C. (1975). *Applied Optimal Control: Optimization, Estimation, and Control* (Revised edition). Hemisphere Publishing Corp., Bristol, PA.
- Conover, W. (1980). *Practical Nonparametric Statistics* (2nd edition). Applied Probability and Statistics Series. John Wiley & Sons, Inc., New York, NY.
- Dennis, J., & Schnabel, R. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Fletcher, R. (1987). *Practical Methods of Optimization* (2nd edition). John Wiley & Sons, Ltd., Chichester, United Kingdom.
- Gibson, J., & Melsa, J. (1975). *Introduction to Nonparametric Detection with Applications*, Vol. 119 of *Mathematics in Science and Engineering*. Academic Press, Inc., New York, NY.

- Gotoh, H., & Yagi, H. (1971). Solid Angle Subtended by a Rectangular Slit. *Nuclear Instruments and Methods*, 96(2), 485–486.
- Hollander, M., & Wolfe, D. (1973). *Nonparametric Statistical Methods*. Applied Probability and Statistics Series. John Wiley & Sons, Inc., New York, NY.
- Jazwinski, A. (1970). *Stochastic Processes and Filter Theory*, Vol. 64 of *Mathematics in Science and Engineering*. Academic Press, Inc., New York, NY.
- Lawrence, C., Zhou, J., & Tits, A. (1994). User's Guide for CFSQP. Tech. rep. TR-94-16r1, Institute for Systems Research, University of Maryland, College Park, MD.
- McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models* (2nd edition)., Vol. 37 of *Statistics and Applied Probability*. Chapman & Hall, Ltd., London, United Kingdom.
- Muske, K., & Rawlings, J. (1995). Nonlinear Moving Horizon State Estimation. In Berber, R. (Ed.), *Methods of Model Based Process Control*, pp. 349–365. Kluwer Academic Publishers, Dordrecht, Netherlands.
- Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical Recipes in C: The Art of Scientific Programming* (2nd edition). Cambridge University Press, Cambridge, United Kingdom.
- Sage, A., & Melsa, J. (1971). *Estimation Theory with Applications to Communications and Control*. Systems Science Series. McGraw-Hill, Inc., New York, NY.
- Tsoufanidis, N. (1983). *Measurement and Detection of Radiation*. Nuclear Engineering Series. McGraw-Hill, Inc., New York, NY.